



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

|   |           |  |
|---|-----------|--|
| <b>(51) International Patent Classification <sup>7</sup> :</b><br><b>C12N 15/11, 15/00, C07K 16/00</b>  | <b>A1</b> | <b>(11) International Publication Number:</b> <b>WO 00/11157</b><br><b>(43) International Publication Date:</b> 2 March 2000 (02.03.00)  |
| <b>(21) International Application Number:</b> PCT/US99/19395<br><b>(22) International Filing Date:</b> 25 August 1999 (25.08.99)<br><b>(30) Priority Data:</b><br>60/097,927 25 August 1998 (25.08.98) US<br><b>(71) Applicant:</b> THE GENERAL HOSPITAL CORPORATION<br>[US/US]; 55 Fruit Street, Boston, MA 02114 (US).<br><b>(72) Inventors:</b> BROWN, Robert, H.; 16 Oakland Avenue, Needham, MA 02192 (US). LIU, Jing; 1629 Van Horn, Outremont, Quebec H2V 1L1 (CA). AOKI, Masashi; Dept. of Neurology, Tohoku University School of Medicine, 1-1, Seiro-Machi, Aoba-ku, Sendai (JP). HO, Meng, F.; Apartment 24, 145 Englewood Avenue, Brighton, MA 02135 (US). MATSUDA-ASADA, Chie; 33 Pond Avenue, Brookline, MA 02445 (US).<br><b>(74) Agent:</b> FRASER, Janis, K.; Fish & Richardson P.C., 225 Franklin Street, Boston, MA 02110-2804 (US). |           | <b>(81) Designated States:</b> CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).<br><b>Published</b><br><i>With international search report.<br/>         Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i> |
| <b>(54) Title:</b> DYSFERLIN, A GENE MUTATED IN DISTAL MYOPATHY AND LIMB GIRDLE MUSCULAR DYSTROPHY<br><div style="text-align: center; margin-top: 20px;"> <p><b>MM candidate region</b></p> </div>  |           |  |
| <b>(57) Abstract</b><br><p>A novel gene and the protein encoded therein, i.e., dysferlin, are disclosed. This gene and its expression products are associated with muscular dystrophy, e.g., Miyoshi myopathy and limb girdle muscular dystrophy 2B.</p>  |           |  |

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

|           |                          |           |  |           |  |           |                          |
|-----------|--------------------------|-----------|--|-----------|--|-----------|--------------------------|
| <b>AL</b> | Albania                  | <b>ES</b> | Spain                                    | <b>LS</b> | Lesotho                                      | <b>SI</b> | Slovenia                 |
| <b>AM</b> | Armenia                  | <b>FI</b> | Finland                                  | <b>LT</b> | Lithuania                                    | <b>SK</b> | Slovakia                 |
| <b>AT</b> | Austria                  | <b>FR</b> | France                                   | <b>LU</b> | Luxembourg                                   | <b>SN</b> | Senegal                  |
| <b>AU</b> | Australia                | <b>GA</b> | Gabon                                    | <b>LV</b> | Latvia                                       | <b>SZ</b> | Swaziland                |
| <b>AZ</b> | Azerbaijan               | <b>GB</b> | United Kingdom                           | <b>MC</b> | Monaco                                       | <b>TD</b> | Chad                     |
| <b>BA</b> | Bosnia and Herzegovina   | <b>GE</b> | Georgia                                  | <b>MD</b> | Republic of Moldova                          | <b>TG</b> | Togo                     |
| <b>BB</b> | Barbados                 | <b>GH</b> | Ghana                                    | <b>MG</b> | Madagascar                                   | <b>TJ</b> | Tajikistan               |
| <b>BE</b> | Belgium                  | <b>GN</b> | Guinea                                   | <b>MK</b> | The former Yugoslav<br>Republic of Macedonia | <b>TM</b> | Turkmenistan             |
| <b>BF</b> | Burkina Faso             | <b>GR</b> | Greece                                   |           |  | <b>TR</b> | Turkey                   |
| <b>BG</b> | Bulgaria                 | <b>HU</b> | Hungary                                  | <b>ML</b> | Mali   | <b>TT</b> | Trinidad and Tobago      |
| <b>BJ</b> | Benin                    | <b>IE</b> | Ireland                                  | <b>MN</b> | Mongolia                                     | <b>UA</b> | Ukraine                  |
| <b>BR</b> | Brazil                   | <b>IL</b> | Israel                                   | <b>MR</b> | Mauritania                                   | <b>UG</b> | Uganda                   |
| <b>BY</b> | Belarus                  | <b>IS</b> | Iceland                                  | <b>MW</b> | Malawi                                       | <b>US</b> | United States of America |
| <b>CA</b> | Canada                   | <b>IT</b> | Italy                                    | <b>MX</b> | Mexico                                       | <b>UZ</b> | Uzbekistan               |
| <b>CF</b> | Central African Republic | <b>JP</b> | Japan                                    | <b>NE</b> | Niger  | <b>VN</b> | Viet Nam                 |
| <b>CG</b> | Congo                    | <b>KE</b> | Kenya                                    | <b>NL</b> | Netherlands                                  | <b>YU</b> | Yugoslavia               |
| <b>CH</b> | Switzerland              | <b>KG</b> | Kyrgyzstan                               | <b>NO</b> | Norway                                       | <b>ZW</b> | Zimbabwe                 |
| <b>CI</b> | Côte d'Ivoire            | <b>KP</b> | Democratic People's<br>Republic of Korea | <b>NZ</b> | New Zealand                                  |           |                          |
| <b>CM</b> | Cameroon                 |           |  | <b>PL</b> | Poland                                       |           |                          |
| <b>CN</b> | China                    | <b>KR</b> | Republic of Korea                        | <b>PT</b> | Portugal                                     |           |                          |
| <b>CU</b> | Cuba                     | <b>KZ</b> | Kazakstan                                | <b>RO</b> | Romania                                      |           |                          |
| <b>CZ</b> | Czech Republic           | <b>LC</b> | Saint Lucia                              | <b>RU</b> | Russian Federation                           |           |                          |
| <b>DE</b> | Germany                  | <b>LI</b> | Liechtenstein                            | <b>SD</b> | Sudan  |           |                          |
| <b>DK</b> | Denmark                  | <b>LK</b> | Sri Lanka                                | <b>SE</b> | Sweden                                       |           |                          |
| <b>EE</b> | Estonia                  | <b>LR</b> | Liberia                                  | <b>SG</b> | Singapore                                    |           |                          |

DYSFERLIN, A GENE MUTATED IN DISTAL MYOPATHY  
AND LIMB GIRDLE MUSCULAR DYSTROPHY

5                   RELATED APPLICATION INFORMATION

This application claims priority from provisional application serial no. 60/097,927, filed August 25, 1998.

Statement as to Federally Sponsored Research

The work described herein was supported in part by  
10 NIH grants 5P01AG12992, 5R01N834913A, and 5P01NS31248.  
The Federal Government therefore may have certain rights  
in the invention.

Background of the Invention

The invention relates to genes involved in the  
15 onset of muscular dystrophy.

Muscular dystrophies constitute a heterogeneous group of disorders. Most are characterized by weakness and atrophy of the proximal muscles, although in rare myopathies such as "Miyoshi myopathy" symptoms may first  
20 arise in distal muscles. Of the various hereditary types of muscular dystrophy, several are caused by mutations or deletions in genes encoding individual components of the dystrophin-associated protein (DAP) complex. It is this DAP complex that links the cytoskeletal protein  
25 dystrophin to the extracellular matrix protein, laminin-2.

Muscular dystrophies may be classified according to the gene mutations that are associated with specific clinical syndromes. For example, mutations in the gene  
30 encoding the cytoskeletal protein dystrophin result in either Duchenne's Muscular Dystrophy or Becker's Muscular Dystrophy, whereas mutations in the gene encoding the extracellular matrix protein merosin produce Congenital

- 2 -

Muscular Dystrophy. Muscular dystrophies with an autosomal recessive mode of inheritance include "Miyoshi myopathy" and the several limb-girdle muscular dystrophies (LGMD2). Of the limb-girdle muscular dystrophies, the deficiencies resulting in LGMD2C, D, E, and F result from mutations in genes encoding the membrane-associated sarcoglycan components of the DAP complex.

#### Summary of the Invention

10 A novel protein, designated dysferlin, is identified and characterized. The dysferlin gene is normally expressed in skeletal muscle cells and is selectively mutated in several families with the hereditary muscular dystrophies, e.g., Miyoshi myopathy  
15 (MM) and limb girdle muscular dystrophy-2B (LGMD2B). These characteristics of dysferlin render it a candidate disease gene for both MM and LGMD2B. An additional novel protein, brain-specific dysferlin, has also been identified. Defects in brain-specific dysferlin may  
20 predispose to selected disorders of the central nervous system. Moreover, the expression of brain-specific dysferlin may be important as a marker for normal neural development (e.g., in vivo or in neural cells in culture). Manipulation of levels of expression of brain-  
25 specific dysferlin, and of the type of expressed brain-specific dysferlin is of use for analyzing the function of brain-specific dysferlin and related dysferlin-associated molecules.

The invention features an isolated DNA which  
30 includes a nucleotide sequence hybridizing under stringent hybridization conditions to a strand of SEQ ID NO:3 or SEQ ID NO:117.



- 3 -

The invention also features an isolated DNA including a nucleotide sequence selected from SEQ ID NOS:4-12.

Also within the invention is an isolated DNA  
5 comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:22-30.

Also within the invention is a single stranded oligonucleotide of 14-50 nucleotides in length having a nucleotide sequence identical to a portion of a strand of  
10 SEQ ID NO:3.

Also within the invention is a pair of PCR primers consisting of:

(a) a first single stranded oligonucleotide consisting of 14-50 contiguous nucleotides of the sense  
15 strand of SEQ ID NO:117; and

(b) a second single stranded oligonucleotide consisting of 14-50 contiguous nucleotides of the antisense strand of SEQ ID NO:117, wherein the sequence of at least one of the oligonucleotides is identical to a  
20 portion of a strand of SEQ ID NO:3, and the first oligonucleotide is not complementary to the second oligonucleotide.

Also within the invention is a pair of single stranded oligonucleotides selected from of SEQ ID NOS  
25 130-231, SEQ ID NO:110, and SEQ ID NO:112.

Also within the invention is an isolated DNA including a nucleotide sequence that encodes a protein that shares at least 70% sequence identity with SEQ ID NO:2, or a complement of the nucleotide sequence.

30 Also within the invention is an isolated DNA including a nucleotide sequence which hybridizes under stringent hybridization conditions to a strand of a nucleic acid, the nucleic acid having a sequence selected from SEQ ID NOS:31-79 and 90-101.

- 4 -

Also within the invention is a single stranded oligonucleotide of 14-50 nucleotides in length having a nucleotide sequence which is identical to a portion of a strand of a nucleic acid selected from SEQ ID NOs:31-79 and 90-100.

Also within the invention is a pair of PCR primers consisting of:

(a) a first single stranded oligonucleotide consisting of 14-50 contiguous nucleotides of the sense strand of a nucleic acid selected from SEQ ID NOs:31-85; and

(b) a second single stranded oligonucleotide consisting of 14-50 contiguous nucleotides of the antisense strand of a nucleic acid selected from SEQ ID NOs:31-85, wherein the sequence of at least one of the oligonucleotides includes a sequence identical to a portion of a strand of a nucleic acid selected from SEQ ID NOs: 31-79 and 90-100, and the first oligonucleotide is not complementary to the second oligonucleotide.

Also within the invention is a pair of single stranded oligonucleotides selected from SEQ ID NOs 101-116, SEQ ID NOs 184-185, SEQ ID NOs 188-191, SEQ ID NOs 210-213, and SEQ ID NOs 216-217.

Also within the invention is a substantially pure protein that has an amino acid sequence sharing at least 70% sequence identity with SEQ ID NO:2.

Also within the invention is a substantially pure protein the sequence of which includes amino acid residues 1-500, 501-1000, 1001-1500, or 1501-2080 of SEQ ID NO:2.

Also within the invention is a substantially pure protein including the amino acid sequence of SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, or SEQ ID NO:89.

In another aspect, the invention features a transgenic non-human mammal having a transgene disrupting

- 5 -

or interfering with the expression of a dysferlin gene, the transgene being chromosomally integrated into the germ cells of the animal.

Another embodiment of the invention features a method of decreasing the symptoms of muscular dystrophy in a mammal by introducing into a cell of the mammal (e.g., a muscle cell or a muscle precursor cell) an isolated DNA which hybridizes under stringent hybridization conditions to a strand of SEQ ID NO:3.

10 Another aspect of the invention provides a method for identifying a patient, a fetus, or a pre-embryo at risk for having a dysferlin-related disorder by (a) providing a sample of genomic DNA from the patient, fetus, or pre-embryo; and (b) determining whether the  
15 sample contains a mutation in a dysferlin gene.

In another aspect, the invention provides a method for identifying a patient, a fetus, or a pre-embryo at risk for having a dysferlin-related disorder by (a) providing a sample including dysferlin mRNA from the  
20 patient, fetus, or pre-embryo; and (b) determining whether the dysferlin mRNA contains a mutation.

Methods of identifying mutations in a dysferlin sequence are useful for predicting (e.g., predicting whether an individual is at risk for developing a  
25 dysferlin-related disorder) or diagnosing disorders associated with dysferlin, e.g., MM and LGMD2B. Such methods can also be used to determine if an individual, fetus, or a pre-embryo is a carrier of a dysferlin mutation, for example in screening procedures. Methods  
30 which distinguish between different dysferlin alleles (e.g., a mutant dysferlin allele and a normal dysferlin allele) can be used to determine carrier status.

The invention also features an isolated nucleic acid comprising a nucleotide sequence which hybridizes  
35 under stringent hybridization conditions to nucleic acids

- 6 -

3284-3720 of SEQ ID NO:232, or the complement of the nucleotide sequence. An isolated nucleic acid including a nucleotide sequence identical to the sequence of nucleotides 3284-3720 of SEQ ID NO:232, or a complement  
5 of the nucleotide sequence is also a feature of the invention. The isolated nucleic acid can include the entire sequence of SEQ ID NO:232 or the complement of SEQ ID NO:232.

Another aspect of the invention features an  
10 isolated polypeptide that includes: a) at least 15 contiguous amino acids of the polypeptide comprising amino acids 1-24 of SEQ ID NO:233, b) a naturally occurring allelic variant of a polypeptide comprising amino acids 1-24 of SEQ ID NO:233, or c) an amino  
15 acid sequence which is encoded by a nucleic acid molecule which hybridizes under stringent conditions to nucleotides 3284-3720 of SEQ ID NO:232. The polypeptide of this aspect can include the entire sequence of SEQ ID NO:233.

20 Also included in the invention is a vector comprising the nucleic acid of claim 44 and a cell that contains the vector. Another aspect of the invention features a method of making a polypeptide by culturing the cell which contains the vector.

25 The invention also features an antibody which specifically binds to a polypeptide of such as those described above. The antibody can bind to a polypeptide selected from amino acids 253-403 of SEQ ID NO:233, amino acids 624-865 of SEQ ID NO:233, and amino acids 1664-1786  
30 of SEQ ID NO:233. Antibodies of the invention can be monoclonal or polyclonal antibodies.

An "isolated DNA" is DNA which has a naturally occurring sequence corresponding to part or all of a given gene but is free of the two genes that normally  
35 flank the given gene in the genome of the organism in

- 7 -

which the given gene naturally occurs. The term therefore includes a recombinant DNA incorporated into a vector, into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote. It also includes a separate molecule such as a cDNA, a genomic fragment, a fragment produced by polymerase chain reaction (PCR), or a restriction fragment, as well as a recombinant nucleotide sequence that is part of a hybrid gene, i.e., a gene encoding a fusion protein. The term excludes intact chromosomes and large genomic segments containing multiple genes contained in vectors or constructs such as cosmids, yeast artificial chromosomes (YACs), and P1-derived artificial chromosome (PAC) contigs.

15 A "noncoding sequence" is a sequence which corresponds to part or all of an intron of a gene, or to a sequence which is 5' or 3' to a coding sequence and so is not normally translated.

An expression control sequence is "operably linked" to a coding sequence when it is within the same nucleic acid and can control expression of the coding sequence.

A "protein" or "polypeptide" is any chain of amino acids linked by peptide bonds, regardless of length or post-translational modification, e.g., glycosylation or phosphorylation.

As used herein, the term "percent sequence identity" means the percentage of identical subunits at corresponding positions in two sequences when the two sequences are aligned to maximize subunit matching, i.e., taking into account gaps and insertions. For purposes of the present invention, percent sequence identity between two polypeptides is to be determined using the Gap program and the default parameters as specified therein.

35 The Gap program is part of the Sequence Analysis Software

- 8 -

Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705.

The algorithm of Myers and Miller, CABIOS (1989) can also be used to determine whether two sequences are similar or identical. Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used.

As used herein, the term "stringent hybridization conditions" means the following DNA hybridization and wash conditions: hybridization at 60°C in the presence of 6 x SSC, 0.5% SDS, 5 x Denhardt's Reagent, and 100 µg/ml denatured salmon sperm DNA; followed by a first wash at room temperature for 20 minutes in 0.5 x SSC and 0.1% SDS and a second wash at 55°C for 30 minutes in 0.2 x SSC and 0.1% SDS.

A "substantially pure protein" is a protein separated from components that naturally accompany it. The protein is considered to be substantially pure when it is at least 60%, by dry weight, free from the proteins and other naturally-occurring organic molecules with which it is naturally associated. Preferably, the purity of the preparation is at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight. A substantially pure dysferlin protein can be obtained, for example, by extraction from a natural source, by expression of a recombinant nucleic acid encoding a dysferlin polypeptide, or by chemical synthesis. Purity can be measured by any appropriate method, e.g., column chromatography, polyacrylamide gel electrophoresis, or HPLC analysis. A chemically synthesized protein or a recombinant protein produced in a cell type other than

- 9 -

the cell type in which it naturally occurs is, by definition, substantially free from components that naturally accompany it. Accordingly, substantially pure proteins include those having sequences derived from eukaryotic organisms but which have been recombinantly produced in *E. coli* or other prokaryotes.

An antibody that "specifically binds" to an antigen is an antibody that recognizes and binds to the antigen, e.g., a dysferlin polypeptide, but which does not substantially recognize and bind to other molecules in a sample (e.g., a biological sample) which naturally includes the antigen, e.g., a dysferlin polypeptide. An antibody that "specifically binds" to dysferlin is sufficient to detect a dysferlin polypeptide in a biological sample using one or more standard immunological techniques (for example, Western blotting or immunoprecipitation).

A "transgene" is any piece of DNA, other than an intact chromosome, which is inserted by artifice into a cell, and becomes part of the genome of the organism which develops from that cell. Such a transgene may include a gene which is partly or entirely heterologous (i.e., foreign) to the host organism, or may represent a gene homologous to an endogenous gene of the organism.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention. The present materials, methods, and examples are illustrative only and not intended to be limiting. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present

- 10 -

specification, including definitions, will control. All the sequences disclosed in the sequence listing are meant to be double-stranded except the sequences of oligonucleotides.

5 Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

#### Brief Description of the Drawings

Fig. 1A is a physical map of the MM locus. Arrows  
10 indicate the five new polymorphic markers and filled, vertical rectangular boxes indicate the previously known polymorphic markers. The five ESTs that are expressed in skeletal muscle are highlighted in bold. Detailed information on the minimal tiling path of the PAC contig  
15 spanning the MM/LGMD2B region is provided in Liu et al., 1998, *Genomics* 49:23-29. The minimal candidate MM region is designated by the solid bracket (top) and compared to the previous candidate region (dashed bracket). TGFA and ADD2 are transforming growth factor alpha and  $\beta$ -adducin  
20 2.

Fig. 1B is a representation of the dysferlin cDNA clones. The probes used in the three successive screens are shown in bold (130347, cDNA10, A27-F2R2). The two most 5' cDNA clones are also shown (B22, B33). The 6.9  
25 kb cDNA for dysferlin (SEQ ID NO:1) is illustrated at the bottom with start and stop codons as shown.

Fig. 1C is a representation of the predicted dysferlin protein. The locations of four C2 domains (SEQ ID NOs: 86-89) are indicated by stippled boxes,  
30 while the putative transmembrane region is hatched. Vertical lines above the cDNA denote the positions of the mutations in Table 2; the associated labels indicate the phenotypes (MM - Miyoshi myopathy; LGMD - limb girdle



- 11 -

muscular dystrophy; DMAT - distal myopathy with anterior tibial onset).

Fig. 2 is the sequence of the predicted 2,080 amino acids of dysferlin (SEQ ID NO:2). The predicted membrane spanning residues are in bold at the carboxy terminus (residues 2047-2063). Partial C2 domains are underlined. Bold, underlined sequences are putative nuclear targeting residues. Possible membrane retention sequences are enclosed within a box.

10 Fig. 3 is a comparison of the Kyle-Doolittle hydrophobicity plots of the dysferlin protein and fer-1. On the Y-axis, increasing positivity corresponds to increasing hydrophobicity. Both proteins have a single, highly hydrophobic stretch at the carboxy terminal end  
15 (arrow). Both share regions of relative hydrophilicity approximately at residue 1,000 (arrowhead).

Fig. 4 is a SSCP analysis of a representative pedigree with dysferlin mutations. Each member of the pedigree is illustrated above the corresponding SSCP  
20 analysis. For each affected individual (solid symbols) shifts are evident in alleles 1 and 2, corresponding respectively to exons 36 and 54. As indicated, the allele 1 and 2 variants are transmitted respectively from the mother and the father. The two affected daughters in  
25 this pedigree have the limb girdle muscular dystrophy (LGMD) phenotype while their affected brother has a pattern of weakness suggestive of Miyoshi myopathy (MM).

Fig. 5 is a representation of the genomic structure of dysferlin. The 55 exons of the dysferlin  
30 gene and their corresponding SEQ ID NOs are indicated below the 6911 bp cDNA (solid line). The cDNA sequences corresponding to SEQ ID NO:1 and SEQ ID NO:3 are shown relative to the 6911 bp cDNA.

Figs. 6A-B are the cDNA sequence of brain-specific  
35 dysferlin (SEQ ID NO:232) and the predicted amino acid

- 12 -

sequence (in single-letter code) of brain-specific dysferlin (SEQ ID NO:233).

#### Detailed Description

The Miyoshi myopathy (MM) locus maps to human  
5 chromosome 2p12-14 between the genetic markers D2S292 and  
D2S286 (Bejaoui et al., 1995, *Neurology* 45:768-72).  
Further refined genetic mapping in MM families placed the  
MM locus between markers GGAA-P7430 and D2S2109 (Bejaoui  
et al., 1998, *Neurogenetics* 1:189-96). Independent  
10 investigation has localized the limb-girdle muscular  
dystrophy (LGMD-2B) to the same genetic interval (Bashir  
et al., 1994, *Hum. Molec. Genetics* 3:455-57; Bashir et  
al., 1996, *Genomics* 33:46-52; Passos-Bueno et al., 1995,  
*Genomics* 27:192-95). Furthermore, two large, inbred  
15 kindreds have been described whose members include both  
MM and LGMD2B patients (Weiler et al., 1996, *Am. J. Hum.*  
*Genet.* 59:872-78; Illarioshkin et al., 1997, *Genomics*  
42:345-48). In these familial studies, the disease  
gene(s) for both MM and LGMD2B mapped to essentially the  
20 same genetic interval. Moreover, in both pedigrees,  
individuals with MM or LGMD2B phenotypes share the same  
haplotypes. This raises the intriguing possibility that  
the two diseases may arise from the same gene defect and  
that a particular disease phenotype is the result of  
25 modification by additional factors.

A 3-Mb PAC contig spanning the entire MM/LGMD2B  
candidate region was recently constructed to facilitate  
the cloning of the MM/LGMD2B gene(s) (Liu et al., 1998,  
*Genomics* 49:23-29). This high resolution PAC contig  
30 resolved the discrepancies of the order of markers in  
previous studies (Bejaoui et al., 1998, *Neurogenetics*  
1:189-96; Bashir et al., 1996, *Genomics* 33:46-52; Hudson  
et al., 1995, *Science* 270:1945-54). The physical size of  
the PAC contig also indicated that the previous minimal

- 13 -

size estimation based on YAC mapping data was significantly underestimated.

#### Identification of Repeat Sequences and Repeat Typing

The PAC contig spanning the MM/LGMD2B region (Liu et al., 1998, *Genomics* 49:23-29) was used as a source for the isolation of new informative markers to narrow the genetic interval of the disease gene(s). DNA from the PAC clones spanning the MM/LGMD2B region was spotted onto Hybond N+™ membrane filters (Amersham, Arlington Heights, IL). The filters were hybridized independently with the following  $\gamma$ -<sup>32</sup>P (Du Pont, Wilmington, DE) labeled repeat sequences: (1) (CA)<sub>15</sub>; (2) pool of (ATT)<sub>10</sub>, (GATA)<sub>8</sub> and (GGAA)<sub>8</sub>; (3) pool of (GAAT)<sub>8</sub>, (GGAT)<sub>8</sub> and (GTAT)<sub>8</sub>; and (4) pool of (AAG)<sub>10</sub> and (ATC)<sub>10</sub>. Hybridization and washing of the filters were carried out at 55°C following standard protocols (Sambrook et al., 1989, *Molecular Cloning: A Laboratory Manual* (2nd Edition), Cold Spring Harbor Press, N.Y.).

Miniprep DNAs of PAC clones containing repeat sequences were digested with restriction enzymes *Hind*III and *Pst*I and ligated into pBluescript II (KS+) vector which is (Stratagene, La Jolla, CA) digested with the same enzymes. Filters of the PAC subclones were hybridized to the  $\gamma$ -<sup>32</sup>P labeled repeats that detected the respective PACs. For clones with an insert size greater than 1 kb the repeat sequences of which could not be identified by a single round of sequencing, the inserts were further subcloned by digestion with *Hae*III and ligation in *Eco*RV-digested pZero-2.1 vector (Invitrogen, Inc., Carlsbad, CA). Miniprep DNAs of the positive subclones were subjected to manual dideoxy sequencing with Sequenase™ enzyme (US Biochemicals, Inc., Cleveland, OH). Primer pairs for amplifying the repeat sequences were selected using the computer program Oligo (Version

- 14 -

4.0, National Biosciences, Inc., Plymouth, MN). Primer sequences are shown in Table 1.

TABLE 1  
New Polymorphic Markers Mapped to the MM/LGMD2B Region

| Marker                 | Repeat  | Primers (5' to 3')  | Annealing<br>T <sub>m</sub> (°C) | Size in<br>PAC (bp) | No. of<br>alleles <sup>1</sup> | Het <sup>2</sup> |
|------------------------|---------|---|----------------------------------|---------------------|--------------------------------|------------------|
| PAC3-H52               | CA      | GATCTAACCTGCTGCTCACC<br>(SEQ ID NO:120)<br>CTGGTGTGTTGCAGAGCGCTG<br>(SEQ ID NO:121) | 57                               | 138                 | 10                             | 0.82             |
| Cy172-H32 <sup>3</sup> | CCAT    | CCTCTCTTCTGCTGCTTCAG<br>(SEQ ID NO:122)<br>TGTGCTGTGTTCCACCTTCGT<br>(SEQ ID NO:123) | 56                               | 199                 | 7                              | 0.72             |
| PAC35-PH2              | CAT     | TCCAAATAGAAATGCCTGAAC<br>(SEQ ID NO:124)<br>AGGTATCACCTCCAAGTGTG<br>(SEQ ID NO:125) | 56                               | 161                 | 5                              | 0.30             |
| PAC16-H41              | Complex | TACCAGCTTCAGAGCTCCCTG<br>(SEQ ID NO:126)<br>TTGATCAGGTGCTCTTGG<br>(SEQ ID NO:127)   | 58                               | 280                 | 4                              | 0.41             |
| Cy7-PH3                | AAGG    | GGAGATTGCTTGAACCCAG<br>(SEQ ID NO:128)<br>TGGCTAATGATGTTGAACATTT<br>(SEQ ID NO:129) | 56                               | 211                 | 4                              | 0.32             |

<sup>1</sup> Observed in 50 unrelated caucasians.

<sup>2</sup> Heterozygosity index.

<sup>3</sup> Located within intron 2 of the *dysferlin* gene.

All oligonucleotides were synthesized by Integrated DNA Technologies, Inc. (Coralville, IA). PCR typing of the repeat markers followed previously described protocols (Bejaoui et al., 1995, Neurology 45:768-772).

- 16 -

Identification of Repeat Markers and Haplotype Analysis

After hybridization with labeled repeat oligos, 17 different groups of overlapping PACs were identified that contained repeat sequences. Some groups contained previously identified repeat markers. For example, five groups of PACs were positively identified by a pool of repeat probes including (ATT)<sub>10</sub>, (GATA)<sub>8</sub>, and (GGAA)<sub>8</sub>. Of these, three groups contained known markers GGAA-P7430 (GGAA repeat), D2S1394 (GATA repeat) and D2S1398 (GGAA repeat) (Hudson et al., 1992, *Nature* 13:622-29; Gastier et al., 1995, *Hum. Molecular Genetics* 4:1829-36). No attempt was made to isolate new repeat markers from these PACs and they were not further analyzed. Similarly, seven groups of PACs that contained known CA repeat markers were excluded. Seven groups of PACs that contained unidentified repeats were retained for further analysis. For each group, the PAC containing the smallest insert was selected for subcloning. Subclones were re-screened and positive clones were sequenced to identify repeats. In total, seven new repeat sequences were identified within the MM/LGMD2B PAC contig. Of these, five are polymorphic within the population that was tested. The information for these five markers is summarized in Table 1. Based on the PAC contig constructed previously across the MM candidate locus (Liu et al., 1998, *Genomics* 48:23-29), the five new markers and ten previously published polymorphic markers were placed in an unambiguous order (Fig. 1).

These markers were analyzed in a large, consanguineous MM family (Bejaoui et al., 1995, *Neurology* 45: 768-72; Bejaoui et al., 1998, *Neurogenetics* 1:189-96). Because MM is a recessive condition, the locus can be defined by identifying regions of the genome that show homozygosity in affected individuals. Conversely, because of the high penetrance of this adult-onset

- 17 -

condition, unaffected adult individuals are not expected to be homozygous by descent across the region. Analysis of haplotype homozygosity in this pedigree indicates that the disease gene lies between markers D2S2111 and PAC3-  
5 H52. Based on the PAC mapping data, the physical distance for this interval is approximately 2.0 Mb. No recombination events were detected between four informative markers (markers cyl172-H32 to PAC16-H41) and the disease locus in family MM-21 (Fig. 1A).

#### 10 Identification of Five Muscle-Expressed ESTs

Twenty-two ESTs and two genes (transforming growth factor alpha [TGF $\alpha$ ] and beta-adducin [ADD2]) were previously mapped to the MM/LGMD2B PAC contig (Fig. 1A) (Liu et al., 1998, *Genomics* 48:23-29). Two  $\mu$ l  
15 (approximately 0.1 ng/ $\mu$ l) of Marathon-ready™ skeletal muscle cDNA (Clontech, Palo Alto, CA) were used as template in a 10  $\mu$ l PCR reaction for analysis of muscle expression of ESTs. The PCR conditions were the same as for the PCR typing of repeat markers. PCR analysis of  
20 skeletal muscle cDNA indicated that five of these ESTs (A006G04, stSG1553R, WI-14958, TIGR-A004Z44 and WI-14051) map within the minimal genetic MM interval of MM and are expressed in skeletal muscle.

Probes were selected corresponding to each of  
25 these five ESTs for Northern blot analysis. cDNA clones (130347, 48106, 172575, 184080, and 510138) corresponding to the five ESTs that are expressed in muscle (respectively TIGR-A004Z44, WI-14051, WI-14958, stSG1553R and A006G04) were selected from the UniGene database  
30 (<http://www.ncbi.nlm.nih.gov/UniGene/>) and obtained from Genome Systems, Inc. (St. Louis, MO). The cDNA probes were first used to screen the MM/LGMD2B PAC filters to confirm that they mapped to the expected position in the MM/LGMD2B contig.

- 18 -

A Northern blot (Clontech) of multiple human tissues was sequentially hybridized to the five cDNA probes and a control  $\beta$ -actin cDNA at 65°C following standard hybridization and washing protocols (Sambrook et al., *supra*). Between hybridizations, probes were removed by boiling the blot at 95-100°C for 4-10 min with 0.5% SDS. The blot was then re-exposed for 24 h to confirm the absence of previous hybridization signals before proceeding with the next round of hybridization.

10       The tissue distribution, intensity of the signals and size of transcripts detected by the five cDNA probes varied. Probes corresponding to ESTs stSG1553R, TIGR-A004Z44 and WI-14958 detected strong signals in skeletal muscle. In addition, the cDNA corresponding to TIGR-  
15 A004Z44 detected a 3.6-3.8 kb brain-specific transcript instead of the 8.5 kb message that was present in other tissues. It is likely that these five ESTs correspond to different genes since the corresponding cDNA probes used for Northern analysis derive from the 3' end of messages,  
20 map to different positions in the MM/LGMD2B contig (Fig. 1A), and differ in their expression patterns.

Current database analysis suggests that three of these ESTs (stSG1553R, WI-14958 and WI-14051) do not match any known proteins (Schuler et al., 1996, Science  
25 274:540-46). A006G04 has weak homology with a protein sequence of unknown function that derives from *C. elegans*. TIGR-A004Z44 has homology only to subdomains present within protein kinase C. Because the five genes corresponding to the ESTs are expressed in skeletal  
30 muscle and map within the minimal genetic interval of the MM/LGMD2B gene(s), they are candidate MM/LGMD2B gene(s).



- 19 -

Cloning of Dysferlin cDNA

EST TIGR-A004Z44 gave a particularly strong skeletal muscle signal on the Northern blot. Moreover, it is bracketed by genetic markers that show no recombination with the disease phenotype in family MM-21 (Fig. 1). The corresponding transcript was therefore cloned and analyzed as a candidate MM gene. From the Unigene database, a cDNA IMAGE clone (130347, 979 bp) was identified that contained the 483 bp EST TIGR-A004Z44.

10 Approximately  $1 \times 10^6$  recombinant clones of a  $\lambda$ gt11 human skeletal muscle cDNA library (Clontech) were plated and screened following standard techniques (Sambrook et al., *supra*). The initial library screening was performed using the insert released from the clone 130347 that  
15 contains EST TIGR-A004Z44, corresponding to the 3' end of the gene. Positive phages were plaque purified and phage DNA was isolated according to standard procedures (Sambrook et al., *supra*). The inserts of the positive clones were released by *EcoRI* digestion of phage DNA and  
20 subsequently subcloned into the *EcoRI* site of pBluescript II (KS+) vector (Stratagene).

Fifty cDNA clones were identified when a human skeletal muscle cDNA library was screened with the 130347 cDNA. Clone cDNA10 with the largest insert (~6.5 kb)  
25 (Fig. 1B) was digested independently with *BamHI* and *PstI* and further subcloned into pBluescript vector. Miniprep DNA of cDNA clones and subclones of cDNA10 was prepared using the Qiagen plasmid Miniprep kit (Valencia, CA). Sequencing was carried out from both ends of each clone  
30 using the SequiTherm EXCEL™ long-read DNA sequencing kit (Epicenter, Madison, WI), fluorescent-labeled M13 forward and reverse primers, and a LI-COR sequencer (Lincoln, NE). Assembly of cDNA contigs and sequence analysis were performed using Sequencher software (Gene Codes  
35 Corporation, Inc., Ann Arbor, MI).

- 20 -

Two additional screens, first with the insert of cDNA10 and then a 683 bp PCR product (A27-F2R2) amplified from the 5' end of the cDNA contig, identified 87 additional cDNA clones. Clones B22 and B33 extended the 5' end by 94 and 20 bp, respectively. The compiled sequence allowed for the generation of a sequence of 6.9 kb (SEQ ID NO:1) (with 10-fold average coverage).

Although the 5' end of the gene has not been further extended to the 8.5 kb predicted by Northern analysis, an open reading frame (ORF) of 6,243 bp has been identified within this 6.9 kb sequence. This ORF is preceded by an in-frame stop codon and begins with the sequence cgcaagcATGCTG (SEQ ID NO:118); five of the first seven bp are consistent with the Kozak consensus sequence for a start codon (Kozak, 1989, *Nucl. Acids Res.* 15:8125-33; Kozak, 1989, *J. Cell. Biol.* 108:229-41). An alternate start codon, in the same frame, +75 bp downstream, appears less likely as a start site GAGACGATGGGG (SEQ ID NO:119). Thus, the entire coding region of this candidate gene is believed to have been identified, as represented by the 6.9 kb sequence contig.

#### Isolation of the Brain-Specific Dysferlin Isoform

##### Identification of the brain-specific isoform of dysferlin

A brain-specific isoform of dysferlin was identified using Northern blot analysis of poly(A+)RNA derived from multiple human adult tissues probed with radiolabeled full-length dysferlin cDNA subclones. A prominent 7.2 kb transcript was detected on Northern blots in skeletal muscle, heart, placenta, lung, and kidney, while a distinct but equally prominent 3.6 kb-3.8 kb transcript was identified exclusively in the brain. Using long exposures, a faint 7.2 kb mRNA was also detected in the

- 21 -

brain. This finding suggested that the shorter brain isoform was likely to be a tissue-specific splice variant of the dysferlin gene. To test this hypothesis, a human brain cDNA library (Stratagene) was screened for the  
5 dysferlin brain isoform.

Cloning of the brain-specific dysferlin isoform

To identify probes that hybridize to the brain-specific dysferlin sequence and so could be used for library screening, fragments of the full-length dysferlin  
10 cDNA clone (derived from a skeletal muscle cDNA library) were generated using restriction enzymes. The fragments were about 1 kb in length and were analyzed by hybridization to a Northern blot that included brain RNA. Sequences suitable for library screening were those that  
15 hybridized to the 3.6-3.8 kb brain-specific transcript. A region of the 3' end of the dysferlin cDNA sequence that is approximately 3 kb in length was identified as hybridizing to brain mRNA. DNA containing sequence from this region was used as a probe for hybridization  
20 screening of a human brain cDNA library (Stratagene).

The human brain cDNA library was plated out and screened using standard procedures. Of the approximately 720,000 plaques screened, 63 primary positive clones were identified. Of these, 20 clones were selected for  
25 further analysis involving standard methods of hybridization, restriction enzyme mapping, and sequencing. The primary positive clones shared regions of overlap with each other.

Sequencing of positive clones, provided 3671  
30 nucleotides of the brain-specific dysferlin sequence (SEQ ID NO:232; Figure 6A-B). The identified sequence corresponds closely to the size of the brain-specific dysferlin transcript detected on Northern blots. With the exception of the 5' region of the sequence, the

- 22 -

brain-specific sequence is identical to about 3.1 kb of the dysferlin sequence (from nucleotide 3722 to 6904 of the dysferlin sequence). In the dysferlin gene, position 3722 corresponds to the start of exon 32. This finding is consistent with the hypothesis that the brain isoform is a splice-variant of the dysferlin gene. At the 5' end of the brain isoform, 489 nucleotides are unique to brain-specific dysferlin. The amino acid sequence encoded by the brain dysferlin nucleic acid sequence (SEQ ID NO:233; Figure 6) contains a unique sequence with an initiation codon within a Kozak consensus sequence. The nucleic acid sequence unique to brain-specific dysferlin encodes a novel 24 amino acid sequence.

#### Identification of Mutations in Miyoshi Myopathy

Two strategies were used to determine whether this 6.9 kb cDNA (SEQ ID NO:1) is mutated in MM. First, the genomic organization of the corresponding gene was determined and the adjoining intronic sequence at each of the 55 exons which make up the cDNA was identified. To identify exon-intron boundaries within the gene, PAC DNA was extracted with the standard Qiagen -Mini Prep protocol. Direct sequencing was performed with DNA Sequence System (Promega, Madison, WI) using <sup>32</sup>P end-labeled primers (Benes et al., 1997, *Biotechniques* 23:98-100). Exon-intron boundaries were identified as the sites where genomic and cDNA sequences diverged. Second, in patients for whom muscle biopsies were available, RT-PCR was also used to prepare cDNA for the candidate gene from the muscle biopsy specimen.

Single strand conformational polymorphism analysis (SSCP) was used to screen each exon in patients from 12 MM families. Putative mutations identified in this way were confirmed by direct sequencing from genomic DNA using exon-specific intronic primers. Approximately 20

- 23 -

ng of total genomic DNA from immortalized lymphocyte cell lines were used as a template for PCR amplification analysis of each exon using primers (below) located in the adjacent introns. SSCP analysis was performed as previously described (Aoki et al., 1998, *Ann. Neurol.* 43:645-53). In patients for whom muscle biopsies were available, mRNA was isolated using RNA-STAT-60™ (Tel-Test, Friendswood, TX) and first-strand cDNA was synthesized from 1-2 µg total RNA with MMLV reverse transcriptase and random hexamer primers (Life Technologies, Gaithersburg, MD). Three µl of this product were used for PCR amplification. Eight sets of primers were designed for muscle cDNA, and overlapping cDNA fragments suitable for SSCP analysis were amplified. After initial denaturation at 94°C for 2 min, amplification was performed using 30 cycles at 94°C for 30 s, 56°C for 30 s, and 72°C for 60 s. The sequences of polymorphisms detected by SSCP analysis were determined by the dideoxy termination method using the Sequenase kit (US Biochemicals). In some instances, the base pair changes predicted corresponding changes in restriction enzyme recognition sites. Such alterations in restriction sites were verified by digesting the relevant PCR products with the appropriate restriction enzymes.

Primer pairs used for SSCP screening and exon sequencing are as follows:

- (1) exon 3, F3261 5'-tctcttctcctagagggccatag-3' (SEQ ID NO: 101) and R326 5'-ctgttcctcccatcgtctcatgg-3' (SEQ ID NO: 102);
- (2) exon 20, F3121 5'-gctcctcccgtgaccctctg-3' (SEQ ID NO: 103) and R3121 5'-gggtcccagccaggagcactg-3' (SEQ ID NO: 104);
- (3) exon 36, F2102 5'-cccctctcaccatctcctgatgtg-3' (SEQ ID NO: 105) and R2111 5'-tggttcaccttcctctacctcgg-3' (SEQ ID NO: 106);

- 24 -

- (4) exon 49, F1081 5'-tcctttggtaggaaatctaggtgg-3'  
(SEQ ID NO: 107) and R1081 5'-ggaagctggacaggcaagagg-3'  
(SEQ ID NO: 108);
- (5) exon 50, F1091 5'-atatactgtgttggaatcttaatgag-3'  
5 (SEQ ID NO: 109) and R1091 5'-gctggcaccacaggggaatcgg-3'  
(SEQ ID NO: 110);
- (6) exon 51, F1101 5'-ctttgcttccttgcataccttctctg-3'  
(SEQ ID NO: 111) and R1101 5'-agcccccatgtgcagaatggg-3'  
(SEQ ID NO: 112);
- 10 (7) exon 52, F1111 5'-ggcagtgatcgagaaacccgg-3' (SEQ  
ID NO: 113) and R1111 5'-catgccctccactggggctgg-3' (SEQ ID  
NO: 114);
- (8) exon 54, F1141 5'-ggatgccagttgactccggg-3' (SEQ ID  
NO: 115) and R1141 5'-ccccaccacagtgtcgtcagg-3' (SEQ ID NO:  
15 116);
- (9) exon 29, F3031 5'-aagtgccaaagcaatgagtgaccgg-3' (SEQ  
ID NO: 184) and R3021 5'-ctcactcccacccaccacctg-3' (SEQ ID  
NO: 185);
- (10) exon 31, F2141 5'-gaatctgccataaccagcttcgtg-3' (SEQ  
20 ID NO: 188) and R2141 5'-tatcaccatagaggcctcgaag-3' (SEQ ID  
NO: 189);
- (11) exon 32, F2981 5'-cagccactcactctggcacctctg-3' (SEQ  
ID NO: 190) and R2981 5'-agcccacagtctctgactctcctg-3' (SEQ ID  
NO: 191);
- 25 (12) exon 43, F2031 5'-cagccaaaccatatcaacaatg-3' (SEQ  
ID NO: 210) and R2021 5'-ctgggggaggtgagggctctag-3' (SEQ ID  
NO: 211);
- (13) exon 44, F2011 5'-gaagtgttttgtctcctcctc-3' (SEQ ID  
NO: 212) and R2011 5'-gcaggcagccagccccatc-3' (SEQ ID NO:  
30 213);
- (14) exon 46, F1041 5'-ctcgtctatgtcttgtgcttgctc-3' (SEQ  
ID NO: 216) and R1051 5'-caccatggtttggggcatgtgg-3' (SEQ ID  
NO: 217).

- 25 -

These primers were used in SSCP screening and exon sequencing, and identified eighteen different mutations in fifteen families (Table 2).

Table 2  
Mutations in Dysferlin in Distal Myopathy and LGMD<sup>1</sup>

| Name      | Nucleotide<br>Change                  | Exon | Consequence          | Origin  | Family<br>name | Allele | Change of<br>restricti<br>on site |
|-----------|---------------------------------------|------|----------------------|---------|----------------|--------|-----------------------------------|
| Mutations |                                       |      |                      |         |                |        |                                   |
| 5 537insA | ins of A<br>at 537                    | 3    | Frameshift           | Arabic  | MM59           | Hom    | no change                         |
| Q605X     | <u>C</u> AG to <u>T</u> AG<br>at 2186 | 20   | Stop at 605          | French  | MM67           | Hom    | -Pst I,<br>-Fnu 4H I <sup>1</sup> |
| I1298V    | <u>A</u> TC to <u>G</u> TC<br>at 4265 | 36   | Amino acid<br>change | Italian | MM,<br>LGMD56  | Het    | -BamHI,<br>-BstYI;<br>+Ava II     |
| E1883X    | <u>G</u> AG to <u>T</u> AG<br>at 5870 | 49   | Stop at<br>1883      | English | MM8            | Het    | no change                         |
| H1857R    | <u>C</u> AT to <u>C</u> GT<br>at 5943 | 50   | Amino acid<br>change | English | MM50           | Het    | no change                         |



|                    |                              |    |                      |          |        |     |                                       |
|--------------------|------------------------------|----|----------------------|----------|--------|-----|---------------------------------------|
| 5966delG           | del of G<br>at 5966          | 50 | Frameshift           | Spanish  | DMAT71 | Hom | no change                             |
| 5966delG           | del of G<br>at 5966          | 50 | Frameshift           | Spanish  | MM75   | Hom | no change                             |
| 6071/6072de<br>LAG | del of AG<br>at<br>6071/6072 | 51 | Frameshift           | English  | MM58   | Het | no change                             |
| 5 6319+1G to<br>A  | Ggt to Gat<br>at 6319+1      | 52 | 5' splice<br>site    | English  | MM8    | Het | no change                             |
| R2042C             | CGT to TGT<br>at 6497        | 54 | Amino acid<br>change | Italian  | MM56   | Het | -Fnu4HI                               |
| R1046H             | CGC to CAG<br>at 3510        | 29 | Amino acid<br>change | Japanese | MM10   | Hom | -HinPI,<br>-Fsp I                     |
| 3746delG           | del of G<br>at 3746          | 31 | Frameshift           | Japanese | MM17   | Hom | -MboII                                |
| 10 Q1160X          | CAG to TAG<br>at 3851        | 32 | Stop at<br>1160      | Mexican  | MM46   | Hom | -ScrFI,<br>-BstNI,<br>+MaeI,<br>+BfaI |

5122/5123de del of CA 43 Frameshift Japanese MM14 Het no change  
 1CA at  
 5122/5123,  
 A to T  
 at 5121

R1586X CGA to TGA 43 Stop at MM12 Hom +Dde I  
 at 5129 1586

5245delG del of G 44 Frameshift French MM63 Hom -Bpm I,  
 at 5245 -BanII 28  
 and G to  
 C at 5249,  
 or G to C  
 at 5245  
 and del G  
 at 5249 + AvaII,  
 +Sau96I

5 E1732X GAG to TAG 46 Stop at MM73 Het -Mbo II  
 at 5567 1732

2573-77 Del of ACCCA at 23 Frameshift Italian MM69  
 Hom ?Please provide  
 del ACCCA 2573-77

<sup>1</sup> MM: Miyoshi myopathy; DMAT: distal myopathy with anterior tibial onset; LGMD: limb girdle muscular dystrophy

<sup>2</sup> +: create a new restriction site, -: eliminate an existing restriction site.

- 30 -

Twelve of the eighteen different mutations are predicted to block dysferlin expression, either through nonsense or frameshift changes. Seven of the thirteen samples are homozygous and thus expected to result in complete loss of dysferlin function. For each mutated exon in these patients, at least 50 control DNA samples (100 chromosomes) were screened to determine the frequencies of the sequence variants. When possible, the parents and siblings of affected individuals were also screened to verify that defined mutations were appropriately co-inherited with the disease in each pedigree (Fig. 4). In two families (50, 58 in Table 2) heterozygous mutations were identified in one allele (respectively a missense mutation and a 2 bp deletion). Mutations in the other allele are presumed to have not been detected (or in three of the screened MM families) either because the mutant and normal SSCP products are indistinguishable or because the mutation lies outside of coding sequence (i.e., in the promoter or a regulatory region of an intron). The disease-associated mutations did not appear to arise in the population as common polymorphisms.

More mutations can be identified by using appropriate primer pairs to amplify an exon and analyze its sequence. The following primer pairs are useful for exon amplification.

| Exon Code | Primer Sequence                                  |
|-----------|--|
| 1 F408    | 5'-gaccacacaagcggcgccctcgg-3' {SEQ ID NO: 130}   |
| F4101     | 5'-gaccccgggcgagggtgggtcgg-3' {SEQ ID NO: 131}   |
| 2 F4111   | 5'-tgtctctccattctccctttttgtg-3' {SEQ ID NO: 132} |
| R4111     | 5'-aggacactgctgagaaggcacctc-3' {SEQ ID NO: 133}  |

- 31 -

|    |             |       |  |
|----|-------------|-------|--|
|    | 3           | F3262 | 5-agtgccttgggtggcacgaagg-3' {SEQ ID    |
|    | NO: 134}    |       |  |
|    |             | R3261 | 5-cctacctgcaccttcaagccatgg-3' {SEQ ID  |
|    | NO: 135}    |       |  |
| 5  | 4           | F3251 | 5-cagaagagccaggggtgccttagg-3' {SEQ ID  |
|    | NO: 136}    |       |  |
|    |             | R3251 | 5-ccttggaccttaacctggcagagg-3' {SEQ ID  |
|    | NO: 137}    |       |  |
|    | 5           | F3242 | 5-cgaggccagcgcaccaacctg-3' {SEQ ID     |
| 10 | NO: 138}    |       |  |
|    |             | R3242 | 5-actgccggccattcttgctggg-3' {SEQ ID    |
|    | NO: 139}    |       |  |
|    | 6           | F3231 | 5-ccaggcctcattagggccctc-3' {SEQ ID     |
|    | NO: 140}    |       |  |
| 15 |             | R3231 | 5-ctgaagaggagcctgggggtcag-3' {SEQ ID   |
|    | NO: 141}    |       |  |
|    | 7           | F3222 | 5-ctgagatttctgactcttgggggtg-3' {SEQ ID |
|    | NO: 142}    |       |  |
|    |             | R3211 | 5-aaggttctgcccctcatgccccatg-3' {SEQ ID |
| 20 | NO: 143}    |       |  |
|    | 8           | F3561 | 5-ctggcctgagggatcagcagg-3' {SEQ ID     |
|    | NO: 144}    |       |  |
|    |             | R3561 | 5-gtgcatacatagcccacggag-3' {SEQ ID     |
|    | NO: 145}    |       |  |
| 25 | 9           | F3551 | 5-gagctattgggttggcgtgtggg-3' {SEQ ID   |
|    | NO: 146}    |       |  |
|    |             | R3552 | 5-accaacacggagaagtgagaactg-3' {SEQ ID  |
|    | NO: 147}    |       |  |
|    | 10          | F3201 | 5-ccacactttattttaacgctttggcgg-3' {SEQ  |
| 30 | ID NO: 148} |       |  |
|    |             | R3201 | 5-cagaaccaaagtgaaggatacgg-3' {SEQ ID   |
|    | NO: 149}    |       |  |
|    | 11          | F3191 | 5-cttctgattctgggatcaccaaagg-3' {SEQ    |
|    | ID NO: 150} |       |  |

- 32 -

|             |       |                                       |
|-------------|-------|---------------------------------------|
|             | F3191 | 5-ggaccgtaaggaagacccaggg-3' {SEQ ID   |
| NO: 151}    |       |                                       |
| 12          | F3181 | 5-cctgtgctcaggagcgcacgaagg-3' {SEQ ID |
| NO: 152}    |       |                                       |
| 5           | R3181 | 5-gcagacctccacccaagggcg-3' {SEQ ID    |
| NO: 153}    |       |                                       |
| 13          | F3171 | 5-gagacagatgggggacagtcaggg-3' {SEQ ID |
| NO: 154}    |       |                                       |
|             | R3171 | 5-cctcccgagagaacctcctg-3' {SEQ ID     |
| 10 NO: 155} |       |                                       |
| 14          | F3161 | 5-gggagcccagagtcacctgg-3' {SEQ ID     |
| NO: 156}    |       |                                       |
|             | R3161 | 5-gggcctccttgggtttgctgg-3' {SEQ ID    |
| NO: 157}    |       |                                       |
| 15          | F3541 | 5-gcctccccagcatcctgccgg-3' {SEQ ID    |
| NO: 158}    |       |                                       |
|             | R3541 | 5-tcactgagccgaatgaaactgagg-3' {SEQ    |
| ID NO: 159} |       |                                       |
| 16          | F3531 | 5-tgtggcctgagttcctttcctgtg-3' {SEQ ID |
| 20 NO: 160} |       |                                       |
|             | R3531 | 5-ggtcaaagggcagaaacgaagagg-3' {SEQ ID |
| NO: 161}    |       |                                       |
| 17          | F3151 | 5-cccgctccttctccagccatg-3' {SEQ ID    |
| NO: 162}    |       |                                       |
| 25          | R3151 | 5-ctcccctgggttgcccccaagg-3' {SEQ ID   |
| NO: 163}    |       |                                       |
| 18          | F3141 | 5-cgaccctctgattgccacttgtg-3' {SEQ ID  |
| NO: 164}    |       |                                       |
|             | R3141 | 5-ggcatacctgcccttgccagg-3' {SEQ ID    |
| 30 NO: 165} |       |                                       |
| 19          | F3522 | 5-tctgtctccctgctccttg-3' {SEQ ID NO:  |
| 166}        |       |                                       |
|             | R3522 | 5-cttccctgccccgacgcccag-3' {SEQ ID    |
| NO: 167}    |       |                                       |

- 33 -

|    |             |       |  |
|----|-------------|-------|--|
|    | 20          | F3121 | 5-gctcctcccgtagaccctctgg-3' {SEQ ID    |
|    | NO: 103}    |       |  |
|    |             | R3121 | 5-gggtcccagccaggagcactg-3' {SEQ ID     |
|    | NO: 104}    |       |  |
| 5  | 21          | F3111 | 5-cagcgctcaggccccgtctctc-3' {SEQ ID    |
|    | NO: 168}    |       |  |
|    |             | R3111 | 5-tgcataggcatgtgcagctttggg-3' {SEQ ID  |
|    | NO: 169}    |       |  |
|    | 22          | F3512 | 5-catgcaccctctgccctgtgg-3' {SEQ ID     |
| 10 | NO: 170}    |       |  |
|    |             | R3512 | 5-agttgagccaggagaggtggg-3' {SEQ ID     |
|    | NO: 171}    |       |  |
|    | 23          | F3101 | 5-catcaggcgcatctccatctgtccg-3' {SEQ ID |
|    | NO: 172}    |       |  |
| 15 |             | R3091 | 5-agcaggagagcagaagaagaaagg-3' {SEQ ID  |
|    | NO: 173}    |       |  |
|    | 24          | F3082 | 5-gtgtgtcaccatccccaccccg-3' {SEQ ID    |
|    | NO: 174}    |       |  |
|    |             | R3082 | 5-caagagatgggagaaaggccttatg-3' {SEQ    |
| 20 | ID NO:175}  |       |  |
|    | 25          | F3073 | 5-ctgggacatccggatcctgaagg-3' {SEQ ID   |
|    | NO: 176}    |       |  |
|    |             | R3073 | 5-tccaggtagtgggaggcagagg-3' {SEQ ID    |
|    | NO: 177}    |       |  |
| 25 | 26          | F3061 | 5-tcccactacctggagctgccttgg-3' {SEQ     |
|    | ID NO: 178} |       |  |
|    |             | R3051 | 5-ggctctccccagccctccctg-3' {SEQ ID     |
|    | NO: 179}    |       |  |
|    | 27          | F3601 | 5-cagagcagcagagactctgaccag-3' {SEQ     |
| 30 | ID NO: 180} |       |  |
|    |             | R3601 | 5-tagaccccacctgccctgag-3' {SEQ ID      |
|    | NO: 181}    |       |  |
|    | 28          | F3501 | 5-tcctctcattgcttgccctgttcgg-3' {SEQ    |
|    | ID NO: 182} |       |  |

- 34 -

|                |       |                                |         |
|----------------|-------|--------------------------------|---------|
|                | R3501 | 5-ttgagagcttgccggggatgg-3'     | {SEQ ID |
| NO: 183}       |       |                                |         |
| 29             | F3031 | 5-aagtgccaaagcaatgagtgaccgg-3' | {SEQ    |
| ID NO: 184}    |       |                                |         |
| 5              | R3021 | 5-ctcactcccacccaccacctg-3'     | {SEQ ID |
| NO: 185}       |       |                                |         |
| 30             | F3011 | 5-cccaccggcctctgagtctgc-3'     | {SEQ ID |
| NO: 186}       |       |                                |         |
|                | R3001 | 5-accctacccaagccaggacaagtg-3'  | {SEQ    |
| 10 ID NO: 187} |       |                                |         |
| 31             | F2141 | 5-gaatctgccataaccagcttcgtg-3'  | {SEQ    |
| ID NO: 188}    |       |                                |         |
|                | R2141 | 5-tatcaccccatagaggcctcgaag-3'  | {SEQ    |
| ID NO: 189}    |       |                                |         |
| 15 32          | F2981 | 5-cagccactcactctggcacctctg-3'  | {SEQ    |
| ID NO: 190}    |       |                                |         |
|                | R2981 | 5-agcccacagtctctgactctcctg-3'  | {SEQ    |
| ID NO: 191}    |       |                                |         |
| 33             | F2131 | 5-acatctctcagggtcctgctgtg-3'   | {SEQ    |
| 20 ID NO: 192} |       |                                |         |
|                | R2211 | 5-cctgtgaggggacgaggcagg-3'     | {SEQ ID |
| NO: 193}       |       |                                |         |
| 34             | F2202 | 5-gccctgggtaagggatgctgattc-3'  | {SEQ    |
| ID NO: 194}    |       |                                |         |
| 25             | R2202 | 5-cctgcctgggcctcctggatc-3'     | {SEQ ID |
| NO: 195}       |       |                                |         |
| 35             | F2111 | 5-gaggggtgatgggggccttagg-3'    | {SEQ ID |
| NO: 196}       |       |                                |         |
|                | R2112 | 5-gcaatcagtttgaagaaggaaagg-3'  | {SEQ    |
| 30 ID NO: 197} |       |                                |         |
| 36             | F2102 | 5-cccctctcaccatctcctgatgtg-3'  | {SEQ    |
| ID NO: 105}    |       |                                |         |
|                | R2111 | 5-ggcttcaccttcctctacctcgg-3'   | {SEQ    |
| ID NO: 106}    |       |                                |         |



- 35 -

|    |             |       |                                     |
|----|-------------|-------|-------------------------------------|
|    | 37          | F2101 | 5-cacctttgtctccattctacctgc-3' {SEQ  |
|    | ID NO: 198} |       |                                     |
|    |             | R2101 | 5-ctcccagccccacgcccagg-3' {SEQ ID   |
|    | NO: 199}    |       |                                     |
| 5  | 38          | F2091 | 5-ctgagccactctcctcattctgtg-3' {SEQ  |
|    | ID NO: 200} |       |                                     |
|    |             | R2091 | 5-tggaaggggacagtagggagg-3' {SEQ ID  |
|    | NO: 201}    |       |                                     |
|    | 39          | F2081 | 5-ggccagtgcgttcttcctcctc-3' {SEQ ID |
| 10 | NO: 202}    |       |                                     |
|    |             | R2071 | 5-tccctgacctgcccacatctc-3' {SEQ ID  |
|    | NO: 203}    |       |                                     |
|    | 40          | F2061 | 5-gccctgtcaggcctggatgg-3' {SEQ ID   |
|    | NO: 204}    |       |                                     |
| 15 |             | R2061 | 5-tgaccagggcctccctggagg-3' {SEQ ID  |
|    | NO: 205}    |       |                                     |
|    | 41          | F2051 | 5-ctgaaatgggtctctttctttctac-3' {SEQ |
|    | ID NO: 206} |       |                                     |
|    |             | R2051 | 5-cacaccgactgtcagactgaagag-3' {SEQ  |
| 20 | ID NO: 207} |       |                                     |
|    | 42          | F2041 | 5-ttgtcccctcctctaatacccatg-3' {SEQ  |
|    | ID NO: 208} |       |                                     |
|    |             | R2041 | 5-ggggttagggacgtcttcgagg-3' {SEQ ID |
|    | NO: 209}    |       |                                     |
| 25 | 43          | F2031 | 5-cagccaaaccatatcaacaatg-3' {SEQ ID |
|    | NO: 210}    |       |                                     |
|    |             | R2021 | 5-ctggggagggtgagggtcttag-3' {SEQ ID |
|    | NO: 211}    |       |                                     |
|    | 44          | F2011 | 5-gaagtgttttgtctcctcctc-3' {SEQ ID  |
| 30 | NO: 212}    |       |                                     |
|    |             | R2011 | 5-gcaggcagccagcccccatc-3' {SEQ ID   |
|    | NO: 213}    |       |                                     |
|    | 45          | F1021 | 5-gggtgccctgtgttggtgac-3' {SEQ ID   |
|    | NO: 214}    |       |                                     |

- 36 -

|                |       |                                |         |
|----------------|-------|--------------------------------|---------|
|                | R1031 | 5-gcaggcagccagcccccata-3'      | {SEQ ID |
| NO: 215}       |       |                                |         |
| 46             | F1041 | 5-ctcgtctatgtcttgtgcttgctc-3'  | {SEQ    |
| ID NO: 216}    |       |                                |         |
| 5              | R1051 | 5-caccatggtttggggcatgtgg-3'    | {SEQ ID |
| NO: 217}       |       |                                |         |
| 47             | F1061 | 5-tctcgcttccccagctcctgc-3'     | {SEQ ID |
| NO: 218}       |       |                                |         |
|                | R1061 | 5-tctggagttcgaggactctggg-3'    | {SEQ ID |
| 10 NO: 219}    |       |                                |         |
| 48             | F1071 | 5-agaaggggtggggagagaacgg-3'    | {SEQ ID |
| NO: 220}       |       |                                |         |
|                | R1071 | 5-cagctcagagcctgtggctgg-3'     | {SEQ ID |
| NO: 221}       |       |                                |         |
| 15 49          | F1082 | 5-aaggccttcccatcctttggtagg-3'  | {SEQ    |
| ID NO: 222}    |       |                                |         |
|                | R1082 | 5-acaaccagagggagcacggg-3'      | {SEQ ID |
| NO: 223}       |       |                                |         |
| 50             | F1092 | 5-gttgacgatgtatatactgtgttg-3'  | {SEQ    |
| 20 ID NO: 224} |       |                                |         |
|                | R1091 | 5-gctggcaccacaggggaatcgg-3'    | {SEQ ID |
| NO: 110}       |       |                                |         |
| 51             | F1102 | 5-gcctctctctaactttgcttccttg-3' | {SEQ    |
| ID NO: 225}    |       |                                |         |
| 25             | R1101 | 5-agcccccatgtgcagaatggg-3'     | {SEQ ID |
| NO: 112}       |       |                                |         |
| 52             | F1112 | 5-ggctacaggctggcagtgatcgag-3'  | {SEQ    |
| ID NO: 226}    |       |                                |         |
|                | R1112 | 5-ttcccccatgccctccactgg-3'     | {SEQ ID |
| 30 NO: 227}    |       |                                |         |
| 53             | F1121 | 5-agccttcgtgcccctaaccaagtg-3'  | {SEQ    |
| ID NO: 228}    |       |                                |         |
|                | R1121 | 5-ctgtgggcattggggctcagg-3'     | {SEQ ID |
| NO: 229}       |       |                                |         |

- 37 -

54 F1141 5-ggatgcccagttgactccggg-3' {SEQ ID  
NO: 115}  
R1141 5-ccccaccacagtgtcgtcagg-3' {SEQ ID  
NO: 116}  
5 55 F1151 5-gccccagtgggatcaccatg-3' {SEQ ID  
NO: 230}  
R116 5-atgctggaggggacccacacgg-3' {SEQ ID  
NO: 231}

### Comparison of Dysferlin With Other Proteins

10 The 6,243 bp ORF of this candidate MM gene is  
predicted to encode 2,080 amino acids (Figs. 1C and 2;  
SEQ ID NO:2). At the amino acid level, this protein is  
highly homologous to the nematode (*Caenorhabditis*  
*elegans*) protein fer-1 (27% identical, 57% identical or  
15 similar: the sequence alignment and comparison was  
performed using [http://vega.igh.cnrs.fr/bin/nph-align\\_query.pl](http://vega.igh.cnrs.fr/bin/nph-align_query.pl).) (Argon & Ward, 1980, *Genetics* 96:413-33;  
Achanzar & Ward, 1997, *J. Cell Science* 110:1073-81).  
This dystrophy-associated, fer-1-like protein has  
20 therefore been designated "dysferlin."

The fer-1 protein was originally identified through  
molecular genetic analysis of a class of fertilization-  
defective *C. elegans* mutants in which spermatogenesis is  
abnormal (Argon & Ward, 1980, *Genetics* 96:413-33). The  
25 mutant fer-1 spermatozoa have defective mobility and show  
imperfect fusion of membranous organelles (Ward et al.,  
1981, *J. Cell Bio.* 91:26-44). Like fer-1, dysferlin is a  
large protein with an extensive, highly charged  
hydrophilic region and a single predicted membrane  
30 spanning region at the carboxy terminus (Fig. 3). There  
is a membrane retention sequence 3' to the membrane  
spanning stretch, indicating that the protein may be  
preferentially targeted to either endoplasmic or  
sarcoplasmic reticulum, probably as a Type II protein

- 38 -

(i.e. with the NH<sub>2</sub> end and most of the following protein located within the cytoplasm) (Fig. 1C). Several nuclear membrane targeting sequences are predicted within the cytoplasmic domain of the protein

5 (<http://psort.nibb.ac.jp/form.html>). Immunocytochemical detection of dysferlin suggests that dysferlin is targeted to or anchored within the sarcoplasmic reticulum.

The cytoplasmic component of this protein contains  
10 four motifs homologous to C2 domains. C2 domains are intracellular protein modules composed of 80 - 130 amino acids (Rizo & Sudhof, 1998, *J. Biol. Chem.* 273:15897). Originally identified within a calcium-dependent isoform of protein kinase C (Nishizuka, 1988, *Nature* 334:661-65),  
15 C2 domains are present in numerous proteins. These domains often arise in approximately homologous pairs described as double C2 or DOC2 domains. One DOC2 protein, DOC2 $\alpha$ , is brain specific and highly concentrated in synaptic vesicles (Orita et al., 1995, *Biochem.*  
20 *Biophys. Res. Comm.* 206:439-48), while another, DOC2 $\beta$ , is ubiquitously expressed (Sakaguchi et al., 1995, *Biochem. Biophys. Res. Comm.* 217:1053-61). Many C2 modules can fold to bind calcium, thereby initiating signaling events such as phospholipid binding. At distal nerve  
25 terminals, for example, the synaptic vesicle protein synaptotagmin has two C2 domains that, upon binding calcium, permit this protein to interact with syntaxin, triggering vesicle fusion with the distal membrane and neurotransmitter release (Sudhof & Rizo, 1996, *Neuron*  
30 17:379-88).

The four dysferlin C2 domains are located at amino acid positions 32-82, 431-475, 1160-1241, and 1582-1660 (Figs. 1C and 3). Indeed, it is almost exclusively through these regions that dysferlin has homology to any  
35 proteins other than fer-1. Each of these segments in

- 39 -

dysferlin is considerably smaller than a typical C2 domain. Moreover, these segments are more widely separated in comparison with the paired C2 regions in synaptotagmin, DOC2 $\alpha$  and  $\beta$  and related C2-positive proteins. For this reason, it is difficult to predict whether the four relatively short C2 domains in dysferlin function analogously to conventional C2 modules. That dysferlin might, by analogy with synaptotagmin, signal events such as membrane fusion is suggested by the fact that fer-1 deficient worms show defective membrane organelle fusion within spermatozoa (Ward et al., 1981, *J. Cell Bio.* 91:26-44).

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

#### EXAMPLES

##### Example 1: Production of dysferlin protein

Standard methods can be used to synthesize either wild type or mutant dysferlin, or fragments of either. These methods can also be used to synthesize brain-specific dysferlin polypeptides including full-length or fragments (e.g., a polypeptide unique to brain-specific dysferlin). For example, a recombinant expression vector encoding dysferlin (or a fragment thereof: e.g., dysferlin minus its membrane-spanning region) operably linked to appropriate expression control sequences can be used to express dysferlin in a prokaryotic (e.g., *E. coli*) or eukaryotic host (e.g., insect cells, yeast cells, or mammalian cells). The protein is then purified by standard techniques. If desired, DNA encoding part or all of the dysferlin sequence can be joined in-frame to DNA encoding a different polypeptide, to produce a chimeric DNA that encodes a hybrid polypeptide. This can be used, for example, to add a tag that will simplify identification or purification of the expressed protein,

- 40 -

or to render the dysferlin (or fragment thereof) more immunogenic.

The preferred means for making short peptide fragments of dysferlin is by chemical synthesis. These  
5 fragments, like dysferlin itself, can be used to generate antibodies, or as positive controls for antibody-based assays.

Fusion proteins are useful, e.g., for generating antibodies. Such fusion proteins are generated using  
10 known methods. In one example, to construct glutathione S-transferase (GST):dysferlin fusion proteins, the BLAST program (Altschul et al., 1990, J. Molec. Biol. 215:403-410) was used to identify three regions of the dysferlin cDNA that show no homology to any known human proteins  
15 (Figure 1). These were subcloned from the dysferlin cDNA as BstYI (881-1333), XmnI (1990-2718) and SalI (5364-5732) fragments ligated respectively into BamHI, SmaI and SalI sites of pGEX-5X-3 (Pharmacia). The three fragments correspond to amino acid sequences at amino acid  
20 locations 253-403, 624-865, and 1664-1786 of SEQ ID NO:2, respectively. The resulting GST fusion proteins of BamHI (43 kDa) and SmaI (53.3 kDa) formed insoluble aggregates that were isolated by SDS-PAGE. The fusion protein of SalI (40.2 kDa) was soluble and thus could be purified  
25 using a glutathione Sepharose 4B column; the SalI dysferlin fragment (14.2 kDa) was isolated by cleavage from GST using Factor Xa protease. The eluted protein was concentrated and further purified by SDS-PAGE. For all three of the fusion peptides, the resulting SDS-PAGE  
30 bands were excised and used to immunize rabbits.

#### Example 2: Production and characterization of anti-dysferlin antibodies

Techniques for generating both monoclonal and polyclonal antibodies specific for a particular protein

- 41 -

are well known. The antibodies can be raised against a short peptide epitope of dysferlin, an epitope linked to a known immunogen to enhance immunogenicity, a long fragment of dysferlin, or the intact protein. Antibodies  
5 can also be raised against brain-specific dysferlin polypeptides, e.g., against amino acids 1-24 of SEQ ID NO:233. Such antibodies raised against dysferlin or brain-specific dysferlin polypeptides are useful for e.g., localizing such polypeptides in tissue sections or  
10 fractionated cell preparations and diagnosing dysferlin-related disorders.

An isolated dysferlin protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind dysferlin using standard techniques  
15 for polyclonal and monoclonal antibody preparation. The dysferlin immunogen can also be a mutant dysferlin or a fragment of a mutant dysferlin. A full-length dysferlin protein can be used or, alternatively, antigenic peptide fragments of dysferlin can be used as immunogens. The  
20 antigenic peptide of dysferlin comprises at least 8 (preferably 10, 15, 20, or 30) amino acid residues of the amino acid sequence shown in SEQ ID NO:2 and encompasses an epitope of such that an antibody raised against the peptide forms a specific immune complex with dysferlin.  
25 Preferred epitopes encompassed by the antigenic peptide are regions of dysferlin that are located on the surface of the protein, e.g., hydrophilic regions.

A dysferlin immunogen typically is used to prepare antibodies by immunizing a suitable subject (e.g.,  
30 rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, recombinantly expressed dysferlin protein or a chemically synthesized dysferlin polypeptide. The preparation can further include an adjuvant, such as  
35 Freund's complete or incomplete adjuvant, or similar

- 42 -

immunostimulatory agent. Immunization of a suitable subject with an immunogenic dysferlin preparation induces a polyclonal anti-dysferlin antibody response.

Polyclonal anti-dysferlin antibodies ("dysferlin  
5 antibodies") can be prepared as described above by immunizing a suitable subject with a dysferlin immunogen. The dysferlin antibody titer in the immunized subject can be monitored over time by standard techniques, such as  
10 with an enzyme linked immunosorbent assay (ELISA) using immobilized dysferlin. If desired, the antibody molecules directed against dysferlin can be isolated from the mammal (e.g., from the blood) and further purified by well-known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after  
15 immunization, e.g., when the dysferlin antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein  
20 (1975) *Nature* 256:495-497, the human B cell hybridoma technique (Kozbor et al. (1983) *Immunol. Today* 4:72), the EBV-hybridoma technique (Cole et al. (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for  
25 producing hybridomas is well known (see generally *Current Protocols in Immunology* (1994) Coligan et al. (eds.) John Wiley & Sons, Inc., New York, NY). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with a  
30 dysferlin immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds dysferlin.

Any of the many well known protocols used for fusing  
35 lymphocytes and immortalized cell lines can be applied



- 43 -

for the purpose of generating a monoclonal antibody against dysferlin (see, e.g., *Current Protocols in Immunology*, supra; Galfre et al. (1977) *Nature* 266:55052; R.H. Kenneth, in *Monoclonal Antibodies: A New Dimension*  
5 *In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); and Lerner (1981) *Yale J. Biol. Med.*, 54:387-402. Moreover, the one in the art will appreciate that there are many variations of such methods which also would be useful. Hybridoma cells producing a  
10 monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind dysferlin, e.g., using a standard ELISA assay.

Alternative to preparing monoclonal antibody-  
15 secreting hybridomas, a monoclonal dysferlin antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (e.g., an antibody phage display library) with dysferlin to thereby isolate immunoglobulin library members that bind dysferlin. Kits  
20 for generating and screening phage display libraries are commercially available (e.g., the Pharmacia Recombinant Phage Antibody System, Catalog No. 27-9400-01; and the Stratagene SurfZAP™ Phage Display Kit, Catalog No. 240612). Additionally, examples of methods and reagents  
25 particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 92/20791; PCT Publication No. WO  
30 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs et al. (1991) *Bio/Technology* 9:1370-1372; Hay et al. (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse et al. (1989) *Science*

- 44 -

246:1275-1281; Griffiths et al. (1993) *EMBO J.* 12:725-734.

As an example, two polyclonal antisera were raised for each of the fusion peptide antigens described above using New Zealand White rabbits. The rabbits were injected with 0.5 mg of antigen using keyhole limpet hemocyanin (KLH) as the adjuvant. Booster injections of 0.25 mg antigen were administered every three weeks over 12 weeks. Serum was prepared from the rabbits and was purified using affinity column chromatography (HiTrap; Pharmacia) or antigen-blotted polyvinylidene difluoride (PVDF) membrane.

Immunoblotting was used to verify that the affinity-purified antisera recognize the cognate fusion peptides by Western immunoblotting (WIB) and that this reactivity was immunoadsorbed by pre-incubation of the antisera with the peptides. Thus, antiserum raised against the polypeptide encoded by the SalI fragment (encoding amino acids 1664-1786) identified the fragment both as a cleaved, 14.2 kDa fragment and as a component of the 40.2 kDa GST-SalI fusion peptide. No reactivity was evident in the fraction containing only the GST fusion partner. Immunoadsorption entirely abolished this staining. Analogous results were detected with all six antisera (to the three different target fusion peptides).

#### Preparation of subcellular fractions

Frozen human muscle (0.3 g) was homogenized in five volumes of 0.25 M sucrose containing proteinase inhibitor (Complete, Boehringer). Subcellular fractions of nuclei, mitochondria, microsomes, and cytosol were separated by differential centrifugation. The purity of each fraction was evaluated by immunoblotting of fraction-specific proteins with antibodies to histone H1 (Calbiochem), cytochrome c (Santa Cruz), Na<sup>+</sup>-K<sup>+</sup> ATPase  $\alpha$ 1 subunit

- 45 -

(Research Diagnostics) and cytosolic superoxide dismutase (Calbiochem).

#### Dysferlin in subcellular fractions

Immunoblotting was used to analyze dysferlin expression. Twenty  $\mu$ g of each subcellular fraction and 40  $\mu$ g of whole homogenate of muscle were separated by SDS-PAGE (4-15% gradient gel) and transferred to a nitrocellulose membrane. Immunoblotting was performed according to standard methods, using chemiluminescence (ECL, Amersham). Immunoblotting of multi-tissue blots identified prominent dysferlin positively at approximately 230 kDa in heart, placenta, skeletal muscle and kidney. Little or no immuno-positive staining was detected in brain, liver, spleen, ovary, or testis. Lower molecular weight bands (approximately 40 kDa) were also evident. Immunoabsorption with the corresponding fusion peptide abolished both the large and the smaller bands. The 230 kDa band was observed with all of the affinity purified, anti-dysferlin antisera.

Immunoblotting of fractionated human muscle documented distinct 230 kDa bands in the whole muscle homogenate and in microsomal and nuclear fractions. Some immunoreactivity was also evident in the nuclear and mitochondrial fractions. No immunoreactivity was detected in the cytosolic fractions. This pattern was seen with all of the anti-dysferlin antisera, and was eliminated by immunoabsorption. The identity of the assayed fractions was verified by Western blotting using fraction-specific antibodies: histone H1 for the nuclear fraction, cytochrome c for the mitochondrial fraction, Na<sup>+</sup>-K<sup>+</sup> ATPase  $\alpha$ 1-subunit for the microsomal fraction, and SOD1 for the cytosolic fraction.

#### Example 3: Diagnosis

- 46 -

The discovery of mutations in the dysferlin gene that are associated with the MM and LMGD2B phenotypes means that individuals can be tested for the disease gene before symptoms appear. This will permit genetic testing and counseling of those with a family history of the disease. Additionally, individuals diagnosed with the genetic defect can be closely monitored for the appearance of symptoms, thereby permitting early intervention, including genetic therapy, as appropriate. Individuals with a brain-specific dysferlin-related disorder can be diagnosed using such methods.

Diagnosis can be carried out on any suitable genomic DNA sample from the individual to be tested. Typically, a blood sample from an adult or child, or a sample of placental or umbilical cord cells of a newborn would be used; alternatively, one could utilize a fetal sample obtained by amniocentesis or chorionic villi sampling.

It is expected that standard genetic diagnostic methods can be used. For example, PCR can be utilized to identify the presence of a deletion, addition, or substitution of one or more nucleotides within any one of the exons of dysferlin. Following the PCR reaction, the PCR product can be analyzed by methods such as a heteroduplex detection technique based upon that of White et al. (1992, *Genomics* 12:301-06), or by techniques such as cleavage of RNA-DNA hybrids using RNase A (Myers et al., 1985, *Science* 230:1242-46), single-stranded conformation polymorphism (SSCP) analysis (Orita et al., 1989, *Genomics* 10:298-99), di-deoxy-fingerprinting (DDF) (Blaszyk et al., 1995, *Biotechniques* 18: 256-260) and denaturing gradient gel electrophoresis (DGGE; Myers et al., 1987, *Methods Enzymol.* 155:501-27). The PCR may be carried out using a primer which adds a G+C rich sequence (termed a "GC-clamp") to one end of the PCR product, thus improving the sensitivity of the subsequent DGGE

- 47 -

procedure (Sheffield et al., 1989, *Proc. Natl. Acad. Sci. USA* 86:232-36). If the particular mutation present in the patient's family is known to have removed or added a restriction site, or to have significantly increased or  
5 decreased the length of a particular restriction fragment, a protocol based upon restriction fragment length polymorphism (RFLP) analysis (perhaps combined with PCR) may be appropriate.

The apparent genetic heterogeneity resulting in the  
10 MM/LGMD2B phenotypes means that the nature of the particular mutation carried by affected individuals in the patient's family may have to be ascertained prior to attempting genetic diagnosis of the patient. Alternatively, a battery of tests designed to identify  
15 any of several mutations known to result in MM/LGMD2B may be utilized to screen individuals without a defined familial genotype. The analysis can be carried out on any genomic DNA derived from the patient, typically from a blood sample.

20 Instead of basing the diagnosis on analysis of the genomic DNA of a patient, one could seek evidence of the mutation in the level or nature of the relevant expression products. Well-known techniques for analyzing expression include mRNA-based methods, such as Northern  
25 blots and *in situ* hybridization (using a nucleic acid probe derived from the relevant cDNA), and quantitative PCR (as described in St-Jacques et al., 1994, *Endocrinology* 134:2645-57). One could also employ polypeptide based methods, including the use of  
30 antibodies specific for the polypeptide of interest. These techniques permit quantitation of the amount of expression of a given gene in the tissue of interest, at least relative to positive and negative controls. One would expect an individual who is heterozygous for a  
35 genetic defect affecting the level of expression of

- 48 -

dysferlin to show up to a 50% loss of expression of this gene in such a hybridization or antibody-based assay. An antibody specific for the carboxy terminal end would be likely to pick up (by failure to bind to) most or all frameshift and premature termination signal mutations, as well as deletions of the carboxy terminal sequence. Use of a battery of monoclonal antibodies specific for different epitopes of dysferlin would be useful for rapidly screening cells to detect those expressing mutant forms of dysferlin (i.e., cells which bind to some dysferlin-specific monoclonal antibodies, but not to others), or for quantifying the level of dysferlin on the surface of cells. One could also use a protein truncation assay (Heim et al., 1994, *Nature Genetics* 8:218-19) to screen for any genetic defect which results in the production of a truncated polypeptide instead of the wild type protein.

Use of immunodetection to identify normal and disease-associated dysferlin

In the following example, immunodetection methods are used to demonstrate a detectable difference in muscles homogenates between normal and disease-associated dysferlin alleles.

Frozen muscle samples (quadriceps) were homogenized in ten volumes of SDS-PAGE sample buffer and boiled for 5 minutes. The final loading volume of SDS-PAGE was adjusted after densitometric measurements (NIH Image) of myosin heavy chain on the Coomassie blue stained gels. Studies were performed on six MM, two LGMD-2B, and three normal muscle samples.

Immunocytochemistry was performed on 8 micron cryostat sections of the muscle that were fixed in 100% cold acetone for 5 minutes and preincubated with PBS containing 1% BSA, 5% heat-inactivated goat serum and 0.2% Triton®X-100. The sections were incubated with

- 49 -

primary antibodies overnight at 4°C and fluorescein-labeled secondary (TAGO Immunologicals) for 30 minutes at room temperature. The primary antibodies were applied in two double staining combinations: SalI-1 anti-dysferlin  
5 and anti-dystrophin antibodies, and SalI-2 anti-dysferlin and anti- $\delta$ -sarcoglycan antibodies. The sections were mounted in SlowFade (Molecular Probes).

The 230 kDA antigen was absent in samples from all five MM patient in immunoblot assays. All five patients  
10 had normal patterns of dystrophin expression. Genetic analysis of the dysferlin gene in the patients predicted that at least two of the five MM patients should have no full-length protein. Two of the other three patients had mutations in at least one allele that are predicted to  
15 eliminate normal dysferlin expression. In all five patients, absence of dysferlin immuno-staining was documented with at least two other anti-dysferlin anti-sera.

Immunostaining of dysferlin, dystrophin and  $\delta$ -  
20 sarcoglycan proteins demonstrated distinct membrane-associated positivity for each protein in normal muscle. By contrast, in both MM and LGMD-2B muscle the dysferlin protein was absent, while the dystrophin and  $\delta$ -sarcoglycan proteins appeared normal.

## 25 Therapeutic Treatment

A patient with MM/LGMD2B, or an individual genetically susceptible to contracting one or both of these diseases, can be treated by supplying dysferlin therapeutic agents of the present invention. Dysferlin  
30 therapeutic agents include a DNA or a subgenomic polynucleotide coding for a functional dysferlin protein. A DNA (e.g., a cDNA) is prepared which encodes the wild type form of the gene operably linked to expression control elements (e.g., promoter and enhancer) that

- 50 -

induce expression in skeletal muscle cells or any other affected cells. The DNA may be incorporated into a vector appropriate for transforming the cells, such as a retrovirus, adenovirus, or adeno-associated virus. One  
5 of the many other known types of techniques for introducing DNA into cells *in vivo* may be used (e.g., liposomes). Particularly useful would be naked DNA techniques, since naked DNA is known to be readily taken up by skeletal muscle cells upon injection into muscle.  
10 Wildtype dysferlin protein can also be administered to an individual who either expresses mutant dysferlin protein or expresses an inadequate amount of dysferlin protein, e.g., a MM/LGMD2B patient.

Administration of the dysferlin therapeutic agents  
15 of the invention can include local or systemic administration, including injection, oral administration, particle gun, or catheterized administration, and topical administration. Various methods can be used to administer the therapeutic dysferlin composition directly  
20 to a specific site in the body. For example, a specific muscle can be located and the therapeutic dysferlin composition injected several times in several different locations within the body of the muscle. The therapeutic dysferlin composition can be directly  
25 administered to the surface of the muscle, for example, by topical application of the composition. X-ray imaging can be used to assist in certain of the above delivery methods. Combination therapeutic agents, including a dysferlin protein or polypeptide or a subgenomic  
30 dysferlin polynucleotide and other therapeutic agents, can be administered simultaneously or sequentially.

Receptor-mediated targeted delivery of therapeutic compositions containing dysferlin subgenomic polynucleotides to specific tissues can also be used.  
35 Receptor-mediated DNA delivery techniques are described



- 51 -

in, for example, Findeis et al. (1993), *Trends in Biotechnol.* 11, 202-05; Chiou et al. (1994), *Gene Therapeutics: Methods and Applications of Direct Gene Transfer* (J.A. Wolff, ed.); Wu & Wu (1988), *J. Biol. Chem.* 263, 621-24; Wu et al. (1994), *J. Biol. Chem.* 269, 542-46; Zenke et al. (1990), *Proc. Natl. Acad. Sci. U.S.A.* 87, 3655-59; Wu et al. (1991), *J. Biol. Chem.* 266, 338-42.

Alternatively, a dysferlin therapeutic composition can be introduced into human cells *ex vivo*, and the cells then implanted into the human. Cells can be removed from a variety of locations including, for example, from a selected muscle. The removed cells can then be contacted with the dysferlin therapeutic composition utilizing any of the above-described techniques, followed by the return of the cells to the human, preferably to or within the vicinity of a muscle. The above-described methods can additionally comprise the steps of depleting fibroblasts or other contaminating non-muscle cells subsequent to removing muscle cells from a human.

Both the dose of the dysferlin composition and the means of administration can be determined based on the specific qualities of the therapeutic composition, the condition, age, and weight of the patient, the progression of the disease, and other relevant factors. If the composition contains dysferlin protein or polypeptide, effective dosages of the composition are in the range of about 1  $\mu$ g to about 100 mg/kg of patient body weight, e.g., about 50  $\mu$ g to about 50 mg/kg of patient body weight, e.g., about 500  $\mu$ g to about 5 mg/kg of patient body weight.

Therapeutic compositions containing dysferlin subgenomic polynucleotides can be administered in a range of about 0.1  $\mu$ g to about 10 mg of DNA/dose for local administration in a gene therapy protocol. Concentration

- 52 -

ranges of about 0.1  $\mu$ g to about 10 mg, e.g., about 1  $\mu$ g to about 1 mg, e.g., about 10  $\mu$ g to about 100  $\mu$ g of DNA can also be used during a gene therapy protocol. Factors such as method of action and efficacy of transformation and expression are considerations that will effect the dosage required for ultimate efficacy of the dysferlin subgenomic polynucleotides. Where greater expression is desired over a larger area of tissue, larger amounts of dysferlin subgenomic polynucleotides or the same amounts readministered in a successive protocol of administrations, or several administrations to different adjacent or close tissue portions of for example, a muscle site, may be required to effect a positive therapeutic outcome. In all cases, routine experimentation in clinical trials will determine specific ranges for optimal therapeutic effect.

#### Animal Model

A line of transgenic animals (e.g., mice, rats, guinea pigs, hamsters, rabbits, or other mammals) can be produced bearing a transgene encoding a defective form of dysferlin. Standard methods of generating such transgenic animals would be used, e.g., as described below.

Alternatively, standard methods of producing null (i.e., knockout) mice could be used to generate a mouse which bears one defective and one wild type allele encoding dysferlin. If desired, two such heterozygous mice could be crossed to produce offspring which are homozygous for the mutant allele. The homozygous mutant offspring would be expected to have a phenotype comparable to the human MM and/or LGMD2B phenotype, and so serve as models for the human disease.

For example, in one embodiment, dysferlin mutations are introduced into a dysferlin gene of a cell, e.g., a

- 53 -

fertilized oocyte or an embryonic stem cell. Such cells can then be used to create non-human transgenic animals in which exogenous altered (e.g., mutated) dysferlin sequences have been introduced into their genome or

5 homologously recombinant animals in which endogenous dysferlin nucleic acid sequences have been altered. Such animals are useful for studying the function and/or activity of dysferlin and for identifying and/or evaluating modulators of dysferlin function. As used

10 herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep,

15 dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene

20 product in one or more cell types or tissues of the transgenic animal. As used herein, an "homologously recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous dysferlin gene has been altered by homologous

25 recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to completed development of the animal.

A transgenic animal of the invention can be created

30 by introducing a nucleic acid encoding a dysferlin mutation into the male pronuclei of a fertilized oocyte, e.g., by microinjection or retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. A dysferlin cDNA sequence e.g., that of

35 (SEQ ID NO:1 or SEQ ID NO:3) can be introduced as a

- 54 -

transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of the human dysferlin gene can be isolated based on hybridization to the human dysferlin sequence (e.g., cDNA) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, U.S. Patent No. 4,873,191 and in Hogan, *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the mutant dysferlin transgene in its genome and/or expression of the mutant dysferlin mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding a mutant dysferlin can further be bred to other transgenic animals carrying other transgenes.

25 To create an homologously recombinant animal, a vector is prepared which contains at least a portion of a dysferlin gene into which a deletion, addition or substitution has been introduced to thereby alter a dysferlin gene. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous dysferlin gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous dysferlin gene is mutated

- 55 -

or otherwise altered (e.g., contains one of the mutations described in Table 2). In the homologous recombination vector, the altered portion of the dysferlin sequence is flanked at its 5' and 3' ends by additional nucleic acid  
5 of the dysferlin gene to allow for homologous recombination to occur between the exogenous dysferlin nucleic acid sequence carried by the vector and an endogenous dysferlin gene in an embryonic stem cell. The additional flanking dysferlin nucleic acid is of  
10 sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see, e.g., Thomas and Capecchi (1987) *Cell* 51:503 for a description of homologous recombination  
15 vectors). The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced dysferlin sequence has homologously recombined with the endogenous dysferlin gene are selected (see, e.g., Li et al. (1992) *Cell* 69:915). The  
20 selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see, e.g., Bradley in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be  
25 implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline  
30 transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley (1991) *Current Opinion in Bio/Technology* 2:823-829 and in PCT Publication Nos. WO 90/11354, WO 91/01140, WO  
35 92/0968, and WO 93/04169.

- 56 -

Other Embodiments

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is  
5 intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

- 57 -

What is claimed is:

1. An isolated DNA comprising a nucleotide sequence which hybridizes under stringent hybridization conditions to SEQ ID NO:3, or a complement thereof.

5        2. The isolated DNA of claim 1, wherein the nucleotide sequence is SEQ ID NO:117.

3. An isolated DNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:4-12.

10       4. The isolated DNA of claim 3, comprising the sequence of SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, or SEQ ID NO:21.

5. An isolated DNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:22-30.

15       6. A single stranded oligonucleotide of 14-50 nucleotides in length having a nucleotide sequence identical to a portion of SEQ ID NO:3, or a complement thereof.

7. A pair of PCR primers consisting of:

20       (a) a first single stranded oligonucleotide consisting of 14-50 contiguous nucleotides that are identical to a portion of SEQ ID NO:117; and

      (b) a second single stranded oligonucleotide consisting of 14-50 contiguous nucleotides that are  
25 identical to a portion of SEQ ID NO:117, wherein the sequence of at least one of the oligonucleotides is identical to a portion of a strand of SEQ ID NO:3, and the first oligonucleotide is not complementary to the second oligonucleotide.

- 58 -

8. A pair of single-stranded oligonucleotides, wherein both oligonucleotides are selected from the group consisting of SEQ ID NOS:130-231, SEQ ID NO:110, and SEQ ID NO:112 and the oligonucleotides are different from  
5 each other.

9. An isolated DNA comprising a nucleotide sequence that encodes a polypeptide that shares at least 70% sequence identity with SEQ ID NO:2, or a complement of the nucleotide sequence.

10 10. The isolated DNA of claim 9, wherein the polypeptide comprises the sequence of SEQ ID NO:2.

11. An isolated DNA comprising a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid having a sequence selected  
15 from the group consisting of SEQ ID NOS:31-79 and 90-100.

12. A single stranded oligonucleotide of 14-50 nucleotides in length comprising a nucleotide sequence which is identical to a portion of a nucleic acid selected from the group consisting of SEQ ID NOS:31-79  
20 and 90-100, or a complement of the nucleotide sequence.

13. The oligonucleotide of claim 12, wherein the portion includes an intronic sequence.

14. A pair of PCR primers consisting of:  
(a) a first single-stranded oligonucleotide  
25 consisting of 14-50 contiguous nucleotides that are identical to a portion of a sense strand of a nucleic acid selected from the group consisting of SEQ ID NOS:31-85; and



- 59 -

(b) a second single stranded oligonucleotide consisting of 14-50 contiguous nucleotides that are identical to a portion of the antisense strand of a nucleic acid selected from the group consisting of SEQ ID  
5 NOS:31-85, wherein the sequence of at least one of the oligonucleotides comprises a sequence identical to a portion of a nucleic acid selected from SEQ ID NOS: 31-79 and 90-100, and wherein the first oligonucleotide is not complementary to the second oligonucleotide.

10        15. A pair of single-stranded oligonucleotides selected from the group consisting of SEQ ID NOS:101-116, SEQ ID NOS:184-185, SEQ ID NOS:188-191, SEQ ID NOS:210-213, and SEQ ID NOS:216-217.

15        16. A vector comprising the isolated DNA of claim 1.

17. A substantially pure polypeptide comprising an amino acid sequence sharing at least 70% sequence identity with SEQ ID NO:2.

20        18. The substantially pure polypeptide of claim 17, wherein the polypeptide comprises an amino acid sequence identical to that of a naturally occurring polypeptide.

19. The substantially pure polypeptide of claim 18, wherein the amino acid sequence comprises the sequence of SEQ ID NO:2.

25        20. A substantially pure polypeptide comprising an amino acid sequence identical to the amino acid sequence of amino acid residues 1-500, 501-1000, 1001-1500, or 1501-2080 of SEQ ID NO:2.

- 60 -

21. A substantially pure polypeptide comprising the amino acid sequence of SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88 or SEQ ID NO:89.

22. A substantially pure polypeptide selected from  
5 the group consisting of amino acids 253-403 of SEQ ID NO:2, amino acids 624-865 of SEQ ID NO:2, and amino acids 1664-1786 of SEQ ID NO:2.

23. A fusion protein comprising a polypeptide of claim 22.

10 24. An antibody that specifically binds to the polypeptide of claim 22.

25. An antibody that binds specifically to the polypeptide of claim 17.

26. A cell comprising the isolated DNA of claim 1.

15 27. A non-human mammal, the genomic DNA of which bears a transgene, wherein the transgene comprises the isolated DNA of claim 1.

28. A transgenic non-human mammal having a transgene disrupting or interfering with the expression  
20 of a dysferlin gene.

29. A method of decreasing the symptoms of muscular dystrophy in a mammal, the method comprising introducing into a cell of said mammal the isolated DNA of claim 1.

30. A method of decreasing the symptoms of muscular  
25 dystrophy in a mammal, the method comprising introducing

- 61 -

into a cell of said mammal the vector of claim 16, the vector being an expression vector.

31. A method of decreasing the symptoms of muscular dystrophy in a mammal, the method comprising introducing  
5 into a cell of said mammal the protein of claim 17.

32. A method for identifying a patient, a fetus, or a pre-embryo at risk for having a dysferlin-related disorder, the method comprising:

(a) obtaining a sample of genomic DNA from the  
10 patient, fetus, or pre-embryo; and

(b) determining whether the sample contains a mutation in a dysferlin gene, wherein a patient, a fetus, or a pre-embryo having a mutation in a dysferlin gene is at risk for having a dysferlin-related disorder.

15 33. The method of claim 32, comprising:

(a) treating the sample of genomic DNA with a restriction enzyme specific for a particular restriction enzyme site; and

(b) detecting the presence or absence of the  
20 particular restriction enzyme site in the sample of genomic DNA as an indication of the presence or absence of a particular mutation in the genomic DNA.

34. The method of claim 33, wherein the restriction enzyme is selected from the group consisting of Pst I,  
25 Fnu4H I, BamH I, BstY I, Ava II, HinP I, Fsp I, Mbo II, ScrF I, BstN I, Mae I, Bfa I, Dde I, Bpm I, Ban II, Ava II, and Sau96 I.

35. The method of claim 32, comprising subjecting the sample to polymerase chain reaction (PCR).

- 62 -

36. The method of claim 32, comprising:

(a) contacting a single stranded oligonucleotide with the sample of genomic DNA; and

(c) detecting hybridization or lack thereof between  
5 the single stranded oligonucleotide and the genomic DNA,  
as an indication of the presence or absence of a mutation  
in the genomic DNA.

37. A method for identifying a patient, a fetus, or  
a pre-embryo at risk for having a dysferlin-related  
10 disorder, said method comprising:

(a) providing a sample comprising dysferlin mRNA  
from the patient, fetus, or pre-embryo; and

(b) determining whether the dysferlin mRNA contains  
a mutation, wherein a patient, a fetus, or a pre-embryo  
15 having a dysferlin mRNA containing a mutation is at risk  
for having a dysferlin-related disorder.

38. The method of claim 37, wherein the presence or  
absence of the mutation is detected by Northern blot.

39. The method of claim 37, wherein the method  
20 includes the step of subjecting the sample to polymerase  
chain reaction (PCR).

40. A method for detecting the absence of a  
mutation in a dysferlin protein of a patient, a fetus, or  
a pre-embryo, the method comprising:

25 (a) providing a sample comprising a dysferlin  
protein of the patient, fetus, or pre-embryo;

(b) contacting the sample with the antibody of  
claim 22; and

(c) detecting binding of the antibody to dysferlin  
30 protein in the sample, if any, wherein binding indicates  
a normal dysferlin protein.

- 63 -

41. An isolated DNA comprising a nucleotide sequence that is identical to the sequence of amino acid residues 3501-3520 of SEQ ID NO:1, 3737-3756 of SEQ ID NO:1, 3842-3861 of SEQ ID NO:1, 5114-5139 of SEQ ID NO:1, or 5239-5255 of SEQ ID NO:1.

42. An isolated DNA comprising a nucleotide sequence selected from the group consisting of  
3501-3520 of SEQ ID NO:1, wherein nucleotide G at 3510 is A;  
3737-3756 of SEQ ID NO:1, wherein nucleotide G at 3746 is deleted;  
3842-3861 of SEQ ID NO:1, wherein nucleotide C at 3851 is T;  
5114-5139 of SEQ ID NO:1, wherein nucleotide C at 5122 and nucleotide A at 5123 are deleted;  
5239-5255 of SEQ ID NO:1, wherein nucleotide G at 5245 is deleted and nucleotide G at 5249 is C; and  
5239-5255 of SEQ ID NO:1, wherein nucleotide G at 5245 is C and nucleotide G at 5249 is deleted.

43. An isolated nucleic acid comprising a nucleotide sequence which hybridizes under stringent hybridization conditions to nucleic acids 3284-3720 of SEQ ID NO:232, or the complement of said nucleotide sequence.

44. An isolated nucleic acid comprising a nucleotide sequence identical to the sequence of nucleotides 3284-3720 of SEQ ID NO:232, or a complement of said nucleotide sequence.

45. The isolated nucleic acid of claim 44, wherein the nucleotide sequence comprises the sequence of SEQ ID NO:232 or the complement of SEQ ID NO:232.

- 64 -

46. An isolated polypeptide comprising:

- a) at least 15 contiguous amino acids of the polypeptide comprising amino acids 1-24 of SEQ ID NO:233,
- b) a naturally occurring allelic variant of a polypeptide comprising amino acids 1-24 of SEQ ID NO:233,  
5 or
- c) an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes under stringent conditions to nucleotides 3284-3720 of SEQ ID NO:232.

10 47. The polypeptide of claim 46, wherein the polypeptide comprises SEQ ID NO:233.

48. A vector comprising the nucleic acid of claim 44.

49. A cell comprising the vector of claim 48.

15 50. A method of making a polypeptide, the method comprising culturing the cell of claim 49.

51. An antibody which specifically binds to a polypeptide of claim 46.

20 52. The antibody of claim 51, wherein the antibody binds to a polypeptide selected from the group comprising amino acids 253-403 of SEQ ID NO:233, amino acids 624-865 of SEQ ID NO:233, and amino acids 1664-1786 of SEQ ID NO:233.

25 53. The antibody of claim 51, wherein the antibody is a monoclonal antibody.

54. The antibody of claim 51, wherein the antibody is a polyclonal antibody.

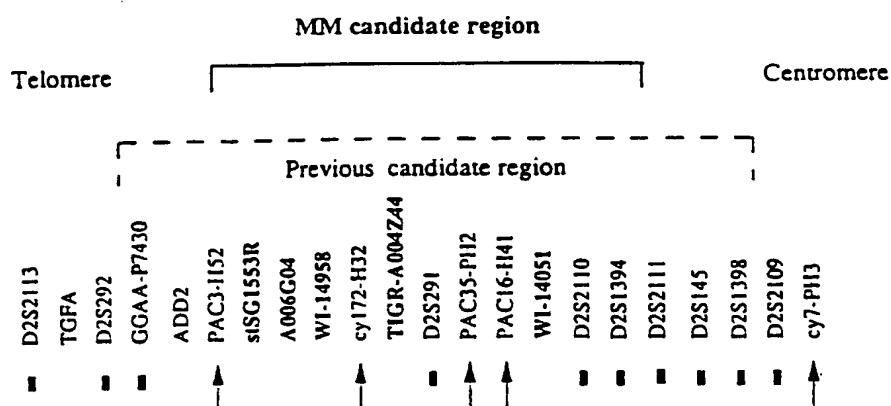


FIG. 1A

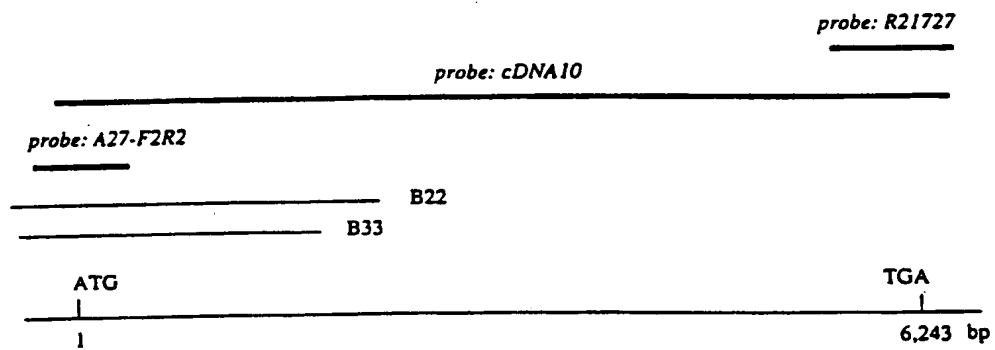


FIG. 1B



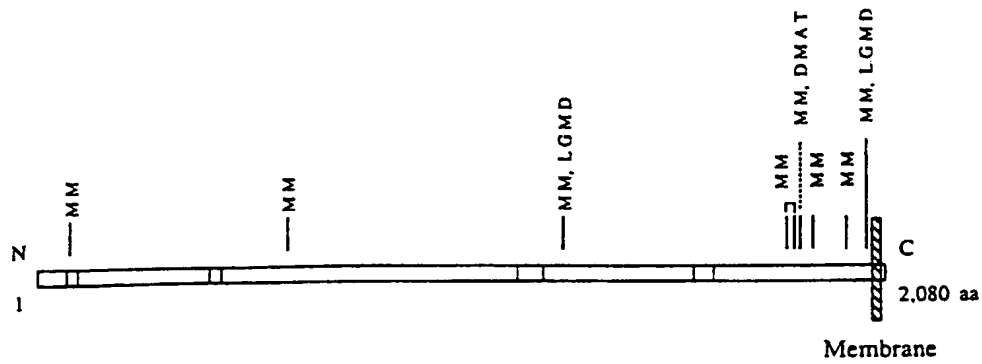


FIG. 1 C

1 ~~MLRVH~~IILYAE NVHTPDTDIS DAYCSAVFAG ~~VKKRTKVIKN~~ ~~SVNPVWNEGF~~  
 51 ~~EWD~~LKGIPLD ~~OGSELHV~~VVK ~~DHETMGRNRF~~ LGEAKVPLRE VLATPSLSAS  
 101 FNAPLLDTKK QPTGASLVQ VSYTPLPGAV PLFPPPTPLE PSPTLPDLVD  
 151 VADTGGEEDT EDQGLTGDEA EPFLDQSGGP GAPTT~~PRKLP~~ ~~SRPPPHYPGI~~  
 201 ~~KKKRS~~SAPTSR KLLSDKPQDF QIRVQVIEGR QLPGVNIKPV VKVTAAGQTK  
 251 RTRIHKGNP LFNETLFFNL FDSPGELFDE PIFITVVDNR SLRTDALLGE  
 301 FRMDVGTIYR EPRHAYLRKW LLLSDPDDFS AGARGYLKTS LCVLGPDEA  
 351 PLERKDPSED KEDIESNLLR PTGVALRGH FCLKVFRAD LPQMDDAVMD  
 401 NVKQIFGFES NKKNLVDPFV EVSFAGKMLC ~~SKILEKTANP~~ ~~OWNONITLPA~~  
 451 ~~MEFSMCEKMR~~ ~~IRIIDWDRLT~~ ~~HNDIVATTYL~~ SMSKISAPGG EIEEEPAGAV  
 501 KPSKASDLDL YLGFLLPTFGP CYINLYGSPR EFTGFDPDYT ELNTGKGEGV  
 551 AYRGRLLLSL ETKLVEHSEQ KVEDLPADDI LRVEKYL~~RRR~~ ~~KYSLFAAFYS~~  
 601 ATMLQDVDDA IQFEVSIGNY GNKFDMTCLP LASTTQYSRA VFDGCHYYL  
 651 PWGNVVPVVV LSSYWEDISH RIETONQLLG IADRLEAGLE QVHLALKAQC  
 701 STEDVDSLVA QLTDELIAGC SSQLGDIHET PSATHLDQYL YQLRTHHLSQ  
 751 ITEAALALKL GHSELPAALE QAEDWLLRLR ALAEEPQNSL PDIWIMLQ  
 801 DKRVAYQRPV AHQVLFSSRG ANYCGKNCCK LQITFLKYPM EKVPGARMPV  
 851 QIRVKLWFGI SVDEKEFNQF AEGKLSVFAE TYENETKLAL VGNWGTGLT  
 901 YPKFSDVTGK IKLPKDSFRP SAGWTWAGDW FVCEKTLH DMDAGHLSFV  
 951 EEVFENQTRL PGGQWIYMSD NYTDVNGEKV LPKDDIECPL GWKWEDEWS  
 1001 TDLNRAVDEQ GWEYSITIPP ~~ERKPKHWVPA~~ ~~EKMYTERRR~~ ~~RWVRLRRRDL~~  
 1051 ~~SOMEAL~~~~KRRR~~ QAEEGEGWE YASLFGWKFH LEYRKTDAPR ~~RRRWRRRMEP~~  
 1101 LEKTGPAAVF ALEGALGGVM DDKSEDSMSV STLSFGVNRP TISCIFYGN  
 1151 RYHLRCYMYQ ~~ARDLAAMDKD~~ ~~SESDPYAIVS~~ ~~FLHOSOKTVV~~ ~~VKNTLNPTWD~~  
 1201 ~~OTLIFYEIEI~~ ~~EGERATVAEO~~ ~~PPSIVVELYD~~ ~~HDTYGADEFM~~ ~~GRCICQPSLE~~  
 1251 RMPRLAWFPL TRGSQPSGEL LASFELIORE KPAIHHPGF EVQETSRILD  
 1301 ESEDIDLPPY PPOREANIYM VPONIKPALQ RTAIEILAWG LRNMKSYQLA  
 1351 NISSPSLVVE CGGQTVQSCV IRNLRKNPNF DICTLFMEVM LPRELYCPP  
 1401 ITVKVIDNRQ FGRRPVVQGC TIRLESFLC DPYSAESPSP QGGPDDVSL  
 1451 SPGEDVLIDI DDKEPLIQ EEFIDWWSK FFASIGEREK CGSYLEKDFD  
 1501 TLKVYDTQLE NVEAFEGLS FCNTFKLYRG KTQEETEDPS VIGEFKGLFK  
 1551 IYPLPEDPAI PMPPRQFHOL AAQGPOECLV ~~RIYIVRAEGL~~ ~~OPKDPNGKCD~~  
 1601 ~~PYIKISIGKK~~ ~~SVSDODNYIP~~ ~~CTLEPVFGKM~~ ~~FELTCTLPLE~~ ~~KDLKITLYDY~~  
 1651 ~~DLISKDEKIG~~ ETVVDLENRL LSKFGARCGP PQTYCVSGPN QWRDQLRPSQ  
 1701 LLHLFCQOHR VKAPVYRTDR VMFQDKEYSI EEIEAGRIPN PHLGPVEERL  
 1751 ALHVLOQOGL VPEHVESRPL YSPLOPDIEQ GKLMWVDFL PKALGRPGPP  
 1801 ~~FNITPERRRR~~ ~~EFLRCIIWNT~~ RDVILDDL SL TGEKMSDIYV KGWMIGFEEH  
 1851 KQKTDVHYRS LGGEGNFNR FIFPFDYLP EQVCTIAKKD AFWRLDKTES  
 1901 KIPARVVFI WDNDKFSFDD FLGSLQLDLN RMPKPAKTAK KCSLDQLDDA  
 1951 FHPEWVSLF EQKTVKGWVP CVAEEGEKKI LAGKLEMTLE IVAESEHEER  
 2001 PAGQGRDEPN MNPKLEDP RR PDSFLWFTS PYKTMKFIW RRRFWAILF  
 2051 IILFILLFL AIFIYAFPNY AAMKLV~~KPBS~~ (SEQ ID NO: 2)

FIG. 2

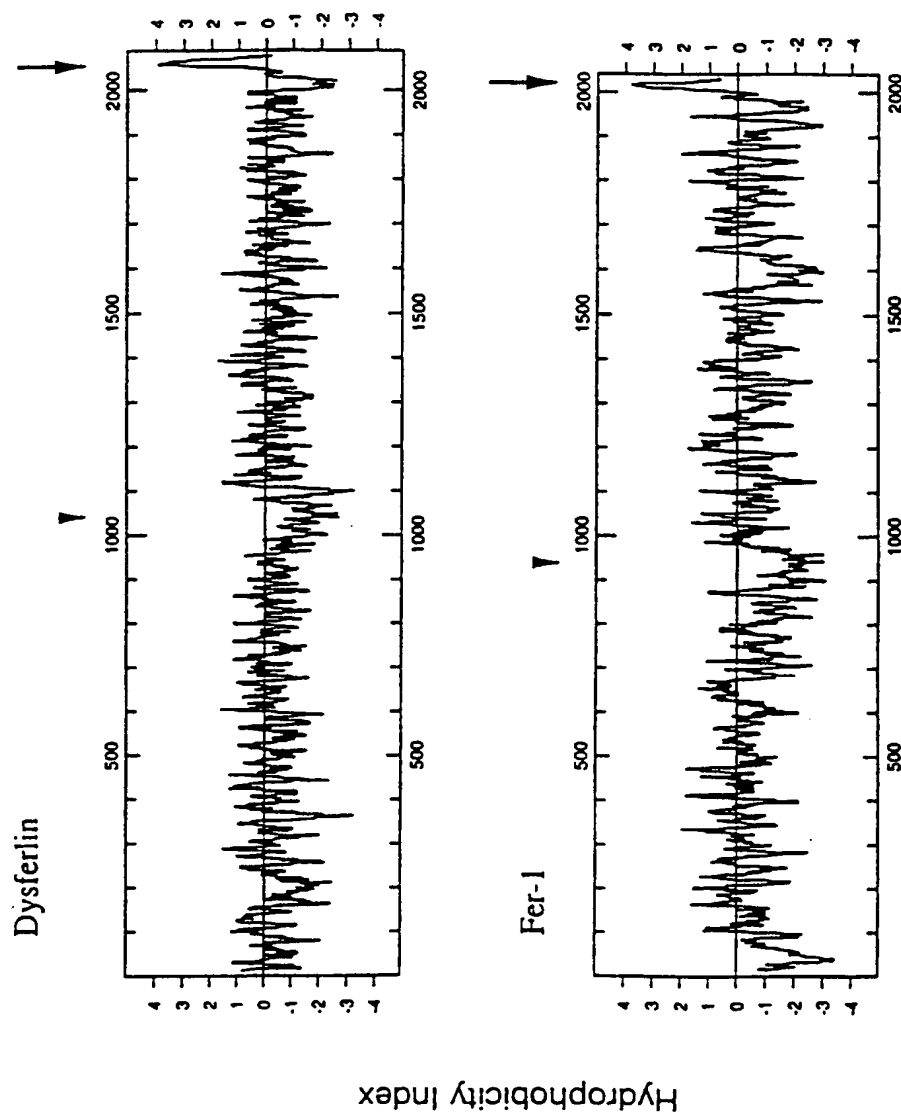


FIG. 3

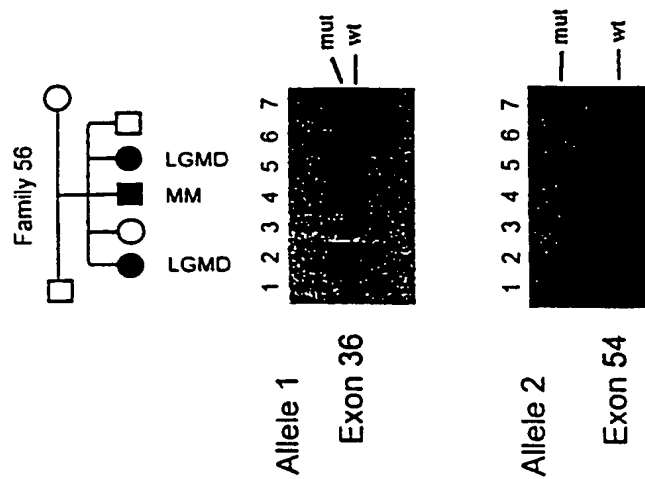


FIG. 4

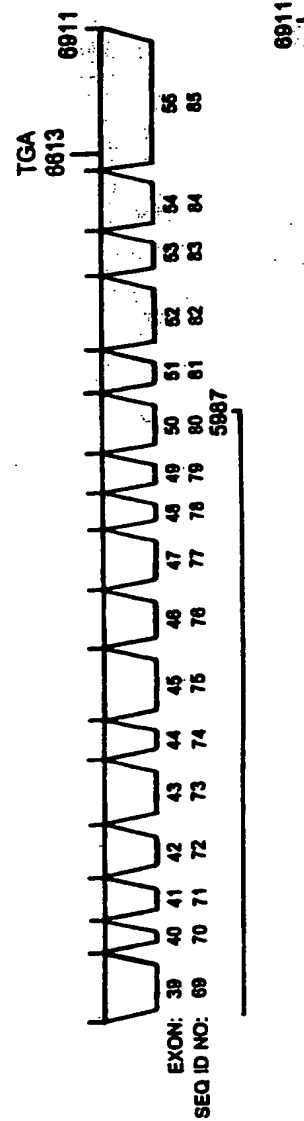
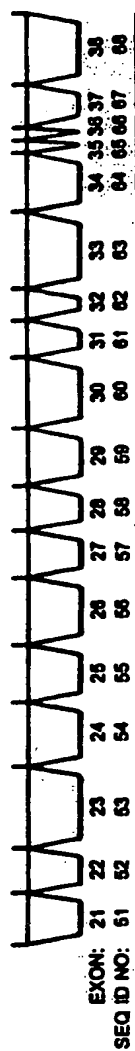
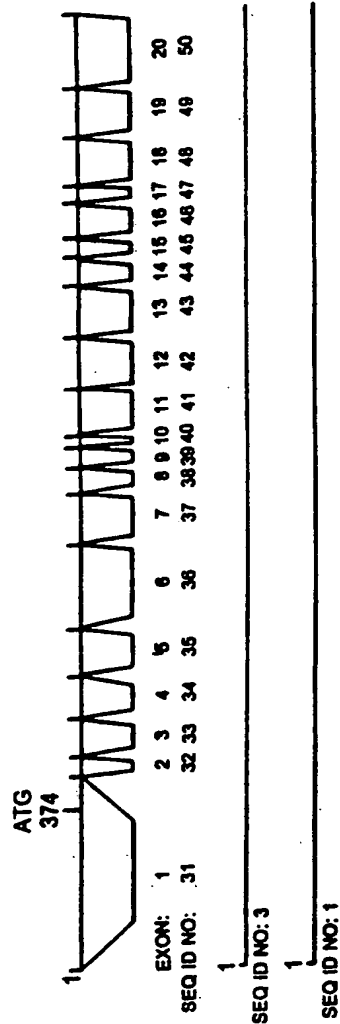


FIG. 5

1/1  
TCC TGG TTC AAG CGA TTC TCT GGC CTC AGC  
S W F K R F S G L S  
91/31  
ATA GAG ACG GGG TTT TGC CAT GTT GGT CAG  
I E T G F C H V G Q  
181/61  
ATT ACA GGC ATG AGT CAC TGT GCC CGG CAG  
I T G M S H C A R Q  
271/91  
GGT ACA AAT AAT TAA TGT AAG CAT AAT CAC  
G T N N \* C K H N H  
361/121  
GGA GAT TGA CTA AGA GGG TGA CCA TCT GGA  
G D \* L R G \* P S G  
451/151  
TTT GTG AAC CCA GGA GGC AGA GGT AGA GAT  
F V N P G R G R D  
541/181  
agc ttc ggt gta aac aga ccc acg att tcc  
S F G V N R P T I S  
631/211  
ctg gct gcg atg gac aag gac tct ttt tct  
L A A M D K D S F S  
721/241  
acc ctt aac ccc acc tgg gac cag acg ctc  
T L N P T W D Q T L  
811/271  
att gtg gtg gac ctg tac gac cat gac act  
I V V E L Y D H D T  
901/301  
cgg ctg gcc tgg ttc cca ctg acg agg ggc  
R L A W F P L T R G  
991/331  
atc cac cat att cct ggt ttt gag gtg cag  
I H H I P G F E V Q  
1081/361  
agg gag gcc aac atc tac atg gtt cct cag  
R E A N I Y M V P Q  
1171/391  
atg aag agt tac cag ctg gcc aac atc tcc  
M K S Y Q L A N I S  
1261/421  
ctc cgg aag aac ccc aac ttt gac atc tgc  
L R K N P N F D I C  
1351/451  
aag gtc atc gat aac cgc cag ttt ggc cgc  
K V I D N R Q F G R  
1441/481  
tcg gcg gag agt cca tcc cca cag ggt ggc  
S A E S P S P Q G G  
1531/511  
gag ccc ctc atc ccc atc cag gag gaa gag  
E P L I P I Q E E E  
1621/541  
tac ctg gag aag gat ttt gac acc ctg aag  
Y L E K D F D T L K  
1711/571  
acc ttc aag ctg tac cgg ggc aag acg cag  
T F K L Y R G K T Q  
1801/601  
ctc cca gaa gac cca gcc atc ccc atg ccc  
L P E D P A I P M P  
1891/631  
att gtc cga gca ttt ggc ctg cag ccc aag  
I V R A F G L Q P K  
1981/661  
gac cag gat aac tac atc ccc tgc acg ctg  
D Q D N Y I P C T L  
2071/691  
aag atc act ctc tat gac tat gac ctc ctc  
K I T L Y D Y D L L  
2161/721  
ttt ggg gct cgc tgt gga ctc cca cag acc  
F G A R C G L P Q T  
2251/751  
ctc ttc tgc cag cag cat aga gtc aag gca  
L F C Q Q H R V K A  
2341/781  
gag gct ggc agg atc cca aac cca cac ctg  
E A G R I P N P H L

31/11  
CTC CCG AGT AGC TGG GAT TAC AGG CAT GCT  
L P S S W D Y R H A  
121/41  
GCT GGT CTC GAA CTC CTG ACC TCA GGT GAT  
A G L E L L T S G D  
211/71  
AGA TGG TCT AAT TCA TAT GAA AGA ACT CTG  
R W S N S Y E R T L  
301/101  
CTA ACC TTG TGG AAT TTT TTT TTT TTT AGA  
L T L W N F F F L R  
391/131  
AAT GAC GTC ATG TGA GAA TGG TTA AAG ATG  
N D V M \* E W L K M  
481/161  
GAG ggc ggc gtg atg gat gac aag  
G E G G V M D D K  
571/191  
tgc ata ttc gac tat ggg aac cgc tac cat  
C I F D Y G N R Y H  
661/221  
gat ccc tat gcc atc gtc tcc ttc ctg cac  
D P Y A I V S F L H  
751/251  
atc ttc tac gag atc gag atc ttt ggc gag  
I F Y E I E I F G E  
841/281  
tat ggt gca gac gag ttt atg ggt cgc tgc  
Y G A D E F M G R C  
931/311  
agc cag ccc tgc ggg gag ctg ctg gcc tct  
S Q P S G E L L A S  
1021/341  
gag aca tca agg atc ctg gat gat tct gag  
E T S R I L D E S E  
1111/371  
aac atc aag cca cgc ctc cag cgt acc gcc  
N I K P A L Q R T A  
1201/401  
tcc ccc agc ctc gtg gta gag tgt ggg ggc  
S P S L V V E C G G  
1291/431  
acc ctc ttc atg gaa gtg atg ctg ccc agg  
T L F M E V M L P R  
1381/461  
cgg cct gtg gtg ggc cag tgt acc atc cgc  
R P V V G Q C T I R  
1471/491  
cca gac gat gtg agc cta ctc agt cct ggg  
P D D V S L L S P G  
1561/521  
ttc atc gat tgg tgg agc aaa ttc ttt gcc  
F I D W W S K F F A  
1651/551  
gtc tat gac aca cag ctg gag aat gtg gag  
V Y D T Q L E N V E  
1741/581  
gag gag aca gaa gat cca tct gtg att ggt  
E E T E D P S V I G  
1831/611  
cca aga cag ttc cac cag ctg gcc gcc cag  
P R Q F H Q L A A Q  
1921/641  
gac ccc aat gga aag tgt gat cct tac atc  
D P N G K C D P Y I  
2011/671  
gag ccc gta ttt gga aag atg ttc gag ctg  
E P V F G K M F E L  
2101/701  
tcc aag gac gaa aag atc ggt gag acg gtc  
S K D E K I G E T V  
2191/731  
tac tgt gtc tct gga cgg aac cag tgg cgg  
Y C V S G P N Q W R  
2281/761  
cct gtg tac cgg aca gac cgt gta atg ttt  
P V Y R T D R V M F  
2371/791  
ggc cca gtg gag gag cgt ctg gct ctg cat  
G P V E E R L A L H

61/21  
CCA CCA AGC CCG GGT AAT TTT GTA TTT TTA  
P P S P G N F V F L  
151/51  
CTG CCC ACC TTG GCC TCC CAA CGT GCT GAG  
L P T L A S Q R A E  
241/81  
AAA AAA GTA GAA AGT GAT TTT CTA AAA TAA  
K K V E S D F L K \*  
331/111  
AGC AAA TTG CAA ATT TGT GAT AGA TCT AAA  
S K L Q I C D R S K  
421/141  
CTC GGG AGA TTG AGC CTA GAG AAA GGA AGA  
L G R L S L E K G R  
511/171  
agt gaa gat tcc atg tcc gtc tcc acc ttg  
S E D S M S V S T L  
601/201  
cta cgc tgc tac atg tac cag gcc cgg gac  
L R C Y M Y Q A R D  
691/231  
cag agc cag aag acg gtg gtg gtg aag aac  
Q S Q K T V V V K N  
781/261  
ccg gcc aca gtt gct gag caa cgg ccc agc  
P A T V A E Q P P S  
871/291  
atc tgt caa cgg agt ctg gaa cgg atg cca  
I C Q P S L E R M P  
961/321  
ttt gag ctc atc cag aga gag aag cgg gcc  
F E L I Q R E K P A  
1051/351  
gac aca gac ctg ccc tac cca cca ccc cag  
D T D L P Y P P P Q  
1141/381  
atc gag atc ctg gca tgg ggc ctg cgg aac  
I E I L A W G L R N  
1231/411  
cag acg ctg cag tcc tgt gtc atc agg aac  
Q T V Q S C V I R N  
1321/441  
gag gag ctc tac tgc ccc ccc atc acc gtc  
E E L Y C P P I T V  
1411/471  
tcc ctg gag agc ttc ctg tgt gac ccc tac  
S L E S F L C D P Y  
1501/501  
gaa gac gtg ctc atc gac att gat gac aag  
E D V L I D I D D K  
1591/531  
tcc ata ggg gag agg gaa aag tgc ggc tcc  
S I G E R E K C G S  
1681/561  
gcc ttt gag ggc ctg tct gac ttt tgt aac  
A F E G L S D F T C N  
1771/591  
gaa ttt aag ggc ctc ttc aaa att tat ccc  
E F K G L F K I Y P  
1861/621  
gga ccc cag gag tgc ttg gtc cgt atc tac  
G P Q E C L V R I Y  
1951/651  
aag atc tcc ata ggg aag aaa tca gtg agt  
K I S I G K K S V S  
2041/681  
acc tgc act ctg cct ctg gag aag gac cta  
T C T L P L E K D L  
2131/711  
gtc gac ctg gag aac agg ctg ctg tcc aag  
V D L E N R L L S K  
2221/741  
gac cag ctc cgc ccc tcc cag ctc ctc cac  
D Q L R P S Q L L H  
2311/771  
cag gat aaa gaa tat tcc att gaa gag ata  
Q D K E Y S I E E I  
2401/801  
gtg ctt cag cag cag ggc ctg gtc cgg gag  
V L Q Q Q G L V P E

Figure 6A

2431/811  
cac gtg gag tca cgg ccc ctc tac tgc ccc  
H V E S R P L Y S P  
2521/841  
ctg ggg cgg cct gga cct ccc ttc aac atc  
L G R P G P P F N I  
2611/871  
atc ctg gat gac ctg agc ctc acg ggg gag  
I L D D L S L T G E  
2701/901  
aca gac gtg cat tat cgt tcc ctg gga ggt  
T D V H Y R S L G G  
2791/931  
tgt acc att gcc aag aag gat gcc ttc tgg  
C T I A K K D A F W  
2881/961  
gac aag ttc tcc ttt gat gat ttt ctg ggc  
D K F S F D D F L G  
2971/991  
ttg gac cag ctg gat gat gct ttc cac cca  
L D Q L D D A F H P  
3061/1021  
gaa gag ggt gag aag aaa ata ctg gcg ggc  
E E G E K K I L A G  
3151/1051  
cag ggc cgg gat gag ccc aac atg aac cct  
Q G R D E P N M N P  
3241/1081  
acc atg aag ttc atc ctg tgg cgg cgt ttc  
T M K F I L W R R F  
3331/1111  
atc tac gcc ttc ccg aac tat gct gcc atg  
I Y A F P N Y A A M  
3421/1141  
cct cca gca tgg gac tgg cct gcc tcc tcc  
P P A W D W P A S S  
3511/1171  
aca gac aga tgg acc ggc cca cac tcc cag  
T D R W T G P H S Q  
3601/1201  
aac gct ttt ttg gat cag ctc aga cat att  
N A F L D Q L R H I  
2461/821  
ctg cag cca gac atc gag cag ggg aag ctg  
L Q P D I E Q G K L  
2551/851  
acc cca cgg aga gcc aga agg ttt ttc ctg  
T P R R A R R F F L  
2641/881  
aag atg agc gac att tat gtg aaa ggt tgg  
K M S D I Y V K G W  
2731/911  
gaa ggc aac ttc aac tgg agg ttc att ttc  
E G N F N W R F I F  
2821/941  
agg ctg gac aag act gag agc aaa atc cca  
R L D K T E S K I P  
2911/971  
tcc ctg cag ctc gat ctc aac cgc atg ccc  
S L Q L D L N R M P  
3001/1001  
gaa tgg ttt gtg tcc ctt ttt gag cag aaa  
E W F V S L F E Q K  
3091/1031  
aag ctg gaa atg acc ttg gag att gta gca  
K L E M T L E I V A  
3181/1061  
aag ctt gag gac cca agg cgc ccc gac acc  
K L E D P R R P D T  
3271/1091  
cgg tgg gcc atc atc ctc ttc atc atc ctc  
R W A I I L F I I L  
3361/1121  
aag ctg gtg aag ccc ttc agc tga gga ctc  
K L V K P F S \* G L  
3451/1151  
gcc cag ctc ggc gag ctc ctc cag acc tcc  
A Q L G E L L Q T S  
3541/1181  
agt tgc taa cat gga gct ctg aga tca ccc  
S C \* H G A L R S P  
3631/1211  
tca gta taa aac agt tgg aac cac aaa aaa  
S V \* N S W N H K K K  
2491  
cag att tgg gtc gac cta ttt ccg aag gcc  
Q M W V D L F P K A  
2581/861  
cgt tgt att atc tgg aat acc aga gat gtg  
R C I I W N T R D V  
2671/891  
atg att ggc ttt gaa gaa cac aag caa aag  
M I G F E E H K Q K  
2761/921  
ccc ttc gac tac ctg cca gct gag caa gtc  
P F D Y L P A E Q V  
2851/951  
gca cga gtg gtg ttc cag atc tgg gac aat  
A R V V F Q I W D N  
2941/981  
aag cca gcc aag aca gcc aag aag tgc tcc  
K P A K T A K K C S  
3031/1011  
aca gtg aag ggc tgg tgg ccc tgt gta gca  
T V K G W W P C V A  
3121/1041  
gag agt gag cat gag gag cgg cct gct ggc  
E S E H E E R P A G  
3211/1071  
tcc ttc ctg tgg ttt acc tcc cca tac aag  
S F L W F T S P Y K  
3301/1101  
ttc atc ctg ctg ctg ttc ctg gcc atc ttc  
F I L L L F L A I F  
3391/1131  
tcc tgc cct gta gaa ggg gcc gtg ggg tcc  
S C P V E G A V G S  
3481/1161  
tag gcc tga ttg tcc tgc cag ggt ggg cag  
\* A \* L S C Q G G Q  
3571/1191  
cac ttc cat cat ttc ctt ctc ccc caa ccc  
H F H H F L L P Q P  
3661/1221  
aaa aaa aaa aa (SEQ ID NO:232)  
(SEQ ID NO:233)

Figure 6B

1/68

## SEQUENCE LISTING

&lt;110&gt; The General Hospital Corporation

&lt;120&gt; DYSFERLIN, A GENE MUTATED IN DISTAL MYOPATHY AND LIMB GIRDLE MUSCULAR DYSTROPHY

&lt;130&gt; 00786/399W02

&lt;150&gt; US 60/097,927

&lt;151&gt; 1998-08-25

&lt;160&gt; 233

&lt;170&gt; FastSEQ for Windows Version 3.0

&lt;210&gt; 1

&lt;211&gt; 6911

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (374)...(6613)

&lt;400&gt; 1

|             |                 |                 |             |             |            |     |
|-------------|-----------------|-----------------|-------------|-------------|------------|-----|
| tcgaccgccc  | agccaggtgc      | aaaatgccgt      | gtcattggga  | gactccgcag  | ccggagcatt | 60  |
| agattacagc  | tcgacggagc      | tcgggaaggg      | cggcgggggt  | ggaagatgag  | cagaagcccc | 120 |
| tgttctcgga  | acgccggctg      | acaagcgggg      | tgagcgcagg  | cggggcgggg  | accagccta  | 180 |
| gcccactgga  | gcagccgggg      | gtggcccgtt      | cccctttaag  | agcaactgct  | ctaagccagg | 240 |
| agccagagat  | tcgagccggc      | ctcgcccagc      | cagccctctc  | cagcgagggg  | accacaagc  | 300 |
| ggcgccctcg  | ccctcccgac      | ctttccgagc      | cctctttgcg  | ccctgggcgc  | acggggccct | 360 |
| acacgcgcca  | agc atg ctg     | agg gtc ttc     | atc ctc tat | gcc gag aac | gtc        | 409 |
|             | Met Leu Arg Val | Phe Ile Leu Tyr | Ala Glu Asn | Val         |            |     |
|             | 1               | 5               | 10          |             |            |     |
| cac aca ccc | gac acc gac     | atc agc gat     | gcc tac tgc | tcc gcg gtg | ttt        | 457 |
| His Thr Pro | Asp Thr Asp     | Ile Ser Asp     | Ala Tyr Cys | Ser Ala Val | Phe        |     |
|             | 15              | 20              | 25          |             |            |     |
| gca ggg gtg | aag aag aga     | acc aaa gtc     | atc aag aac | agc gtg aac | cct        | 505 |
| Ala Gly Val | Lys Lys Arg     | Thr Lys Val     | Ile Lys Asn | Ser Val Asn | Pro        |     |
|             | 30              | 35              | 40          |             |            |     |
| gta tgg aat | gag gga ttt     | gaa tgg gac     | ctc aag ggc | atc ccc ctg | gac        | 553 |
| Val Trp Asn | Glu Gly Phe     | Glu Trp Asp     | Leu Lys Gly | Ile Pro Leu | Asp        |     |
|             | 45              | 50              | 55          | 60          |            |     |
| cag ggc tct | gag ctt cat     | gtg gtg gtc     | aaa gac cat | gag acg atg | ggg        | 601 |
| Gln Gly Ser | Glu Leu His     | Val Val Val     | Lys Asp His | Glu Thr Met | Gly        |     |
|             | 65              | 70              | 75          |             |            |     |
| agg aac agg | ttc ctg ggg     | gaa gcc aag     | gtc cca ctc | cga gag gtc | ctc        | 649 |
| Arg Asn Arg | Phe Leu Gly     | Glu Ala Lys     | Val Pro Leu | Arg Glu Val | Leu        |     |
|             | 80              | 85              | 90          |             |            |     |
| gcc acc cct | agt ctg tcc     | gcc agc ttc     | aat gcc ccc | ctg ctg gac | acc        | 697 |
| Ala Thr Pro | Ser Leu Ser     | Ala Ser Phe     | Asn Ala Pro | Leu Leu Asp | Thr        |     |
|             | 95              | 100             | 105         |             |            |     |
| aag aag cag | ccc aca ggg     | gcc tcg ctg     | gtc ctg cag | gtg tcc tac | aca        | 745 |
| Lys Lys Gln | Pro Thr Gly     | Ala Ser Leu     | Val Leu Gln | Val Ser Tyr | Thr        |     |
|             | 110             | 115             | 120         |             |            |     |



2/68

|   |      |
|---|------|
| ccg ctg cct gga gct gtg ccc ctg ttc ccg ccc cct act cct ctg gag<br>Pro Leu Pro Gly Ala Val Pro Leu Phe Pro Pro Pro Thr Pro Leu Glu<br>125 130 135 140 | 793  |
| ccc tcc ccg act ctg cct gac ctg gat gta gtg gca gac aca gga gga<br>Pro Ser Pro Thr Leu Pro Asp Leu Asp Val Val Ala Asp Thr Gly Gly<br>145 150 155     | 841  |
| gag gaa gac aca gag gac cag gga ctc act gga gat gag gcg gag cca<br>Glu Glu Asp Thr Glu Asp Gln Gly Leu Thr Gly Asp Glu Ala Glu Pro<br>160 165 170     | 889  |
| ttc ctg gat caa agc gga ggc ccg ggg gct ccc acc acc cca agg aaa<br>Phe Leu Asp Gln Ser Gly Gly Pro Gly Ala Pro Thr Thr Pro Arg Lys<br>175 180 185     | 937  |
| cta cct tca cgt cct ccg ccc cac tac ccc ggg atc aaa aga aag cga<br>Leu Pro Ser Arg Pro Pro Pro His Tyr Pro Gly Ile Lys Arg Lys Arg<br>190 195 200     | 985  |
| agt gcg cct aca tct aga aag ctg ctg tca gac aaa ccg cag gat ttc<br>Ser Ala Pro Thr Ser Arg Lys Leu Leu Ser Asp Lys Pro Gln Asp Phe<br>205 210 215 220 | 1033 |
| cag atc agg gtc cag gtg atc gag ggg cgc cag ctg ccg ggg gtg aac<br>Gln Ile Arg Val Gln Val Ile Glu Gly Arg Gln Leu Pro Gly Val Asn<br>225 230 235     | 1081 |
| atc aag cct gtg gtc aag gtt acc gct gca ggg cag acc aag cgg acg<br>Ile Lys Pro Val Val Lys Val Thr Ala Ala Gly Gln Thr Lys Arg Thr<br>240 245 250     | 1129 |
| cgg atc cac aag gga aac agc cca ctc ttc aat gag act ctt ttc ttc<br>Arg Ile His Lys Gly Asn Ser Pro Leu Phe Asn Glu Thr Thr Phe Phe<br>255 260 265     | 1177 |
| aac ttg ttt gac tct cct ggg gag ctg ttt gat gag ccc atc ttt atc<br>Asn Leu Phe Asp Ser Pro Gly Glu Leu Phe Asp Glu Pro Ile Phe Ile<br>270 275 280     | 1225 |
| acg gtg gta gac tct cgt tct ctc agg aca gat gct ctc ctc ggg gag<br>Thr Val Val Asp Ser Arg Ser Leu Arg Thr Asp Ala Leu Leu Gly Glu<br>285 290 295 300 | 1273 |
| ttc cgg atg gac gtg ggc acc att tac aga gag ccc ccg cac gcc tat<br>Phe Arg Met Asp Val Gly Thr Ile Tyr Arg Glu Pro Arg His Ala Tyr<br>305 310 315     | 1321 |
| ctc agg aag tgg ctg ctg ctc tca gac cct gat gac ttc tct gct ggg<br>Leu Arg Lys Trp Leu Leu Leu Ser Asp Pro Asp Asp Phe Ser Ala Gly<br>320 325 330     | 1369 |
| gcc aga ggc tac ctg aaa aca agc ctt tgt gtg ctg ggg cct ggg gac<br>Ala Arg Gly Tyr Leu Lys Thr Ser Leu Cys Val Leu Gly Pro Gly Asp<br>335 340 345     | 1417 |
| gaa gcg cct ctg gag aga aaa gac ccc tct gaa gac aag gag gac att<br>Glu Ala Pro Leu Glu Arg Lys Asp Pro Ser Glu Asp Lys Glu Asp Ile<br>350 355 360     | 1465 |
| gaa agc aac ctg ctc cgg ccc aca ggc gta gcc ctg cga gga gcc cac<br>Glu Ser Asn Leu Leu Arg Pro Thr Gly Val Ala Leu Arg Gly Ala His<br>365 370 375 380 | 1513 |
| ttc tgc ctg aag gtc ttc cgg gcc gag gac ttg ccg cag atg gac gat<br>Phe Cys Leu Lys Val Phe Arg Ala Glu Asp Leu Pro Gln Met Asp Asp<br>385 390 395     | 1561 |

3/68

|   |      |
|---|------|
| gcc gtg atg gac aac gtg aaa cag atc ttt ggc ttc gag agt aac aag<br>Ala Val Met Asp Asn Val Lys Gln Ile Phe Gly Phe Glu Ser Asn Lys<br>400 405 410     | 1609 |
| aag aac ttg gtg gac ccc ttt gtg gag gtc agc ttt gcg ggg aaa atg<br>Lys Asn Leu Val Asp Pro Phe Val Glu Val Ser Phe Ala Gly Lys Met<br>415 420 425     | 1657 |
| ctg tgc agc aag atc ttg gag aag acg gcc aac cct cag tgg aac cag<br>Leu Cys Ser Lys Ile Leu Glu Lys Thr Ala Asn Pro Gln Trp Asn Gln<br>430 435 440     | 1705 |
| aac atc aca ctg cct gcc atg ttt ccc tcc atg tgc gaa aaa atg agg<br>Asn Ile Thr Leu Pro Ala Met Phe Pro Ser Met Cys Glu Lys Met Arg<br>445 450 455 460 | 1753 |
| att cgt atc ata gac tgg gac cgc ctg act cac aat gac atc gtg gct<br>Ile Arg Ile Ile Asp Trp Asp Arg Leu Thr His Asn Asp Ile Val Ala<br>465 470 475     | 1801 |
| acc acc tac ctg agt atg tgc aaa atc tct gcc cct gga gga gaa ata<br>Thr Thr Tyr Leu Ser Met Ser Lys Ile Ser Ala Pro Gly Gly Glu Ile<br>480 485 490     | 1849 |
| gaa gag gag cct gca ggt gct gtc aag cct tgc aaa gcc tca gac ttg<br>Glu Glu Glu Pro Ala Gly Ala Val Lys Pro Ser Lys Ala Ser Asp Leu<br>495 500 505     | 1897 |
| gat gac tac ctg ggc ttc ctc ccc act ttt ggg ccc tgc tac atc aac<br>Asp Asp Tyr Leu Gly Phe Leu Pro Thr Phe Gly Pro Cys Tyr Ile Asn<br>510 515 520     | 1945 |
| ctc tat ggc agt ccc aga gag ttc aca ggc ttc cca gac ccc tac aca<br>Leu Tyr Gly Ser Pro Arg Glu Phe Thr Gly Phe Pro Asp Pro Tyr Thr<br>525 530 535 540 | 1993 |
| gag ctc aac aca ggc aag ggg gaa ggt gtg gct tat cgt ggc cgg ctt<br>Glu Leu Asn Thr Gly Lys Gly Glu Gly Val Ala Tyr Arg Gly Arg Leu<br>545 550 555     | 2041 |
| ctg ctc tcc ctg gag acc aag ctg gtg gag cac agt gaa cag aag gtg<br>Leu Leu Ser Leu Glu Thr Lys Leu Val Glu His Ser Glu Gln Lys Val<br>560 565 570     | 2089 |
| gag gac ctt cct gcg gat gac atc ctc cgg gtg gag aag tac ctt agg<br>Glu Asp Leu Pro Ala Asp Asp Ile Leu Arg Val Glu Lys Tyr Leu Arg<br>575 580 585     | 2137 |
| agg cgc aag tac tcc ctg ttt gcg gcc ttc tac tca gcc acc atg ctg<br>Arg Arg Lys Tyr Ser Leu Phe Ala Ala Phe Tyr Ser Ala Thr Met Leu<br>590 595 600     | 2185 |
| cag gat gtg gat gat gcc atc cag ttt gag gtc agc atc ggg aac tac<br>Gln Asp Val Asp Asp Ala Ile Gln Phe Glu Val Ser Ile Gly Asn Tyr<br>605 610 615 620 | 2233 |
| ggg aac aag ttc gac atg acc tgc ctg ccg ctg gcc tcc acc act cag<br>Gly Asn Lys Phe Asp Met Thr Cys Leu Pro Leu Ala Ser Thr Thr Gln<br>625 630 635     | 2281 |
| tac agc cgt gca gtc ttt gac ggg tgc cac tac tac tac cta ccc tgg<br>Tyr Ser Arg Ala Val Phe Asp Gly Cys His Tyr Tyr Tyr Leu Pro Trp<br>640 645 650     | 2329 |

4/68

|   |      |
|---|------|
| ggt aac gtg aaa cct gtg gtg gtg ctg tca tcc tac tgg gag gac atc<br>Gly Asn Val Lys Pro Val Val Val Leu Ser Ser Tyr Trp Glu Asp Ile<br>655 660 665     | 2377 |
| agc cat aga atc gag act cag aac cag ctg ctt ggg att gct gac cgg<br>Ser His Arg Ile Glu Thr Gln Asn Gln Leu Leu Gly Ile Ala Asp Arg<br>670 675 680     | 2425 |
| ctg gaa gct ggc ctg gag cag gtc cac ctg gcc ctg aag gcg cag tgc<br>Leu Glu Ala Gly Leu Glu Gln Val His Leu Ala Leu Lys Ala Gln Cys<br>685 690 695 700 | 2473 |
| tcc acg gag gac gtg gac tcg ctg gtg gct cag ctg acg gat gag ctc<br>Ser Thr Glu Asp Val Asp Ser Leu Val Ala Gln Leu Thr Asp Glu Leu<br>705 710 715     | 2521 |
| atc gca ggc tgc agc cag cct ctg ggt gac atc cat gag aca ccc tct<br>Ile Ala Gly Cys Ser Gln Pro Leu Gly Asp Ile His Glu Thr Pro Ser<br>720 725 730     | 2569 |
| gcc acc cac ctg gac cag tac ctg tac cag ctg cgc acc cat cac ctg<br>Ala Thr His Leu Asp Gln Tyr Leu Tyr Gln Leu Arg Thr His His Leu<br>735 740 745     | 2617 |
| agc caa atc act gag gct gcc ctg gcc ctg aag ctc ggc cac agt gag<br>Ser Gln Ile Thr Glu Ala Ala Leu Ala Leu Lys Leu Gly His Ser Glu<br>750 755 760     | 2665 |
| ctc cct gca gct ctg gag cag gcg gag gac tgg ctc ctg cgt ctg cgt<br>Leu Pro Ala Ala Leu Glu Gln Ala Glu Asp Trp Leu Leu Arg Leu Arg<br>765 770 775 780 | 2713 |
| gcc ctg gca gag gag ccc cag aac agc ctg ccg gac atc gtc atc tgg<br>Ala Leu Ala Glu Glu Pro Gln Asn Ser Leu Pro Asp Ile Val Ile Trp<br>785 790 795     | 2761 |
| atg ctg cag gga gac aag cgt gtg gca tac cag cgg gtg ccc gcc cac<br>Met Leu Gln Gly Asp Lys Arg Val Ala Tyr Gln Arg Val Pro Ala His<br>800 805 810     | 2809 |
| caa gtc ctc ttc tcc cgg cgg ggt gcc aac tac tgt ggc aag aat tgt<br>Gln Val Leu Phe Ser Arg Arg Gly Ala Asn Tyr Cys Gly Lys Asn Cys<br>815 820 825     | 2857 |
| ggg aag cta cag aca atc ttt ctg aaa tat ccg atg gag aag gtg cct<br>Gly Lys Leu Gln Thr Ile Phe Leu Lys Tyr Pro Met Glu Lys Val Pro<br>830 835 840     | 2905 |
| ggc gcc cgg atg cca gtg cag ata cgg gtc aag ctg tgg ttt ggg ctc<br>Gly Ala Arg Met Pro Val Gln Ile Arg Val Lys Leu Trp Phe Gly Leu<br>845 850 855 860 | 2953 |
| tct gtg gat gag aag gag ttc aac cag ttt gct gag ggg aag ctg tct<br>Ser Val Asp Glu Lys Phe Asn Gln Phe Ala Glu Gly Lys Leu Ser<br>865 870 875         | 3001 |
| gtc ttt gct gaa acc tat gag aac gag act aag ttg gcc ctt gtt ggg<br>Val Phe Ala Glu Thr Tyr Glu Asn Glu Thr Lys Leu Ala Leu Val Gly<br>880 885 890     | 3049 |
| aac tgg ggc aca acg ggc ctc acc tac ccc aag ttt tct gac gtc acg<br>Asn Trp Gly Thr Thr Gly Leu Thr Tyr Pro Lys Phe Ser Asp Val Thr<br>895 900 905     | 3097 |
| ggc aag atc aag cta ccc aag gac agc ttc cgc ccc tcg gcc ggc tgg<br>Gly Lys Ile Lys Leu Pro Lys Asp Ser Phe Arg Pro Ser Ala Gly Trp<br>910 915 920     | 3145 |

5/68

|   |      |
|---|------|
| acc tgg gct gga gat tgg ttc gtg tgt ccg gag aag act ctg ctc cat<br>Thr Trp Ala Gly Asp Trp Phe Val Cys Pro Glu Lys Thr Leu Leu His<br>925 930 935 940     | 3193 |
| gac atg gac gcc ggt cac ctg agc ttc gtg gaa gag gtg ttt gag aac<br>Asp Met Asp Ala Gly His Leu Ser Phe Val Glu Glu Val Phe Glu Asn<br>945 950 955         | 3241 |
| cag acc cgg ctt ccc gga ggc cag tgg atc tac atg agt gac aac tac<br>Gln Thr Arg Leu Pro Gly Gly Gln Trp Ile Tyr Met Ser Asp Asn Tyr<br>960 965 970         | 3289 |
| acc gat gtg aac ggg gag aag gtg ctt ccc aag gat gac att gag tgc<br>Thr Asp Val Asn Gly Glu Lys Val Leu Pro Lys Asp Ile Glu Cys<br>975 980 985             | 3337 |
| cca ctg ggc tgg aag tgg gaa gat gag gaa tgg tcc aca gac ctc aac<br>Pro Leu Gly Trp Lys Trp Glu Asp Glu Glu Trp Ser Thr Asp Leu Asn<br>990 995 1000        | 3385 |
| cgg gct gtc gat gag caa ggc tgg gag tat agc atc acc atc ccc ccg<br>Arg Ala Val Asp Glu Gln Gly Trp Glu Tyr Ser Ile Thr Ile Pro Pro<br>1005 1010 1015 1020 | 3433 |
| gag cgg aag ccg aag cac tgg gtc cct gct gag aag atg tac tac aca<br>Glu Arg Lys Pro Lys His Trp Val Pro Ala Glu Lys Met Tyr Tyr Thr<br>1025 1030 1035      | 3481 |
| cac cga cgg cgg cgc tgg gtg cgc ctg cgc agg agg gat ctc agc caa<br>His Arg Arg Arg Arg Trp Val Arg Leu Arg Arg Arg Asp Leu Ser Gln<br>1040 1045 1050      | 3529 |
| atg gaa gca ctg aaa agg cac agg cag gcg gag gcg gag ggc gag ggc<br>Met Glu Ala Leu Lys Arg His Arg Gln Ala Glu Ala Glu Gly Glu Gly<br>1055 1060 1065      | 3577 |
| tgg gag tac gcc tct ctt ttt ggc tgg aag ttc cac ctc gag tac cgc<br>Trp Glu Tyr Ala Ser Leu Phe Gly Trp Lys Phe His Leu Glu Tyr Arg<br>1070 1075 1080      | 3625 |
| aag aca gat gcc ttc cgc cgc cgc cgc tgg cgc cgt cgc atg gag cca<br>Lys Thr Asp Ala Phe Arg Arg Arg Arg Trp Arg Arg Arg Met Glu Pro<br>1085 1090 1095 1100 | 3673 |
| ctg gag aag acg ggg cct gca gct gtg ttt gcc ctt gag ggg gcc ctg<br>Leu Glu Lys Thr Gly Pro Ala Ala Val Phe Ala Leu Glu Gly Ala Leu<br>1105 1110 1115      | 3721 |
| ggc ggc gtg atg gat gac aag agt gaa gat tcc atg tcc gtc tcc acc<br>Gly Gly Val Met Asp Asp Lys Ser Glu Asp Ser Met Ser Val Ser Thr<br>1120 1125 1130      | 3769 |
| ttg agc ttc ggt gtg aac aga ccc acg att tcc tgc ata ttc gac tat<br>Leu Ser Phe Gly Val Asn Arg Pro Thr Ile Ser Cys Ile Phe Asp Tyr<br>1135 1140 1145      | 3817 |
| ggg aac cgc tac cat cta cgc tgc tac atg tac cag gcc cgg gac ctg<br>Gly Asn Arg Tyr His Leu Arg Cys Tyr Met Tyr Gln Ala Arg Asp Leu<br>1150 1155 1160      | 3865 |
| gct gcg atg gac aag gac tct ttt tct gat ccc tat gcc atc gtc tcc<br>Ala Ala Met Asp Lys Asp Ser Phe Ser Asp Pro Tyr Ala Ile Val Ser<br>1165 1170 1175 1180 | 3913 |

6/68

|   |      |
|---|------|
| ttc ctg cac cag agc cag aag acg gtg gtg gtg aag aac acc ctt aac<br>Phe Leu His Gln Ser Gln Lys Thr Val Val Val Lys Asn Thr Leu Asn<br>1185 1190 1195      | 3961 |
| ccc acc tgg gac cag acg ctc atc ttc tac gag atc gag atc ttt ggc<br>Pro Thr Trp Asp Gln Thr Leu Ile Phe Tyr Glu Ile Glu Ile Phe Gly<br>1200 1205 1210      | 4009 |
| gag ccg gcc aca gtt gct gag caa ccg ccc agc att gtg gtg gag ctg<br>Glu Pro Ala Thr Val Ala Glu Gln Pro Pro Ser Ile Val Val Glu Leu<br>1215 1220 1225      | 4057 |
| tac gac cat gac act tat ggt gca gac gag ttt atg ggt cgc tgc atc<br>Tyr Asp His Asp Thr Tyr Gly Ala Asp Glu Phe Met Gly Arg Cys Ile<br>1230 1235 1240      | 4105 |
| tgt caa ccg agt ctg gaa ccg atg cca ccg ctg gcc tgg ttc cca ctg<br>Cys Gln Pro Ser Leu Glu Arg Met Pro Arg Leu Ala Trp Phe Pro Leu<br>1245 1250 1255 1260 | 4153 |
| acg agg ggc agc cag ccg tcc ggg gag ctg ctg gcc tct ttt gag ctc<br>Thr Arg Gly Ser Gln Pro Ser Gly Glu Leu Leu Ala Ser Phe Glu Leu<br>1265 1270 1275      | 4201 |
| atc cag aga gag aag ccg gcc atc cac cat att cct ggt ttt gag gtg<br>Ile Gln Arg Glu Lys Pro Ala Ile His His Ile Pro Gly Phe Glu Val<br>1280 1285 1290      | 4249 |
| cag gag aca tca agg atc ctg gat gag tct gag gac aca gac ctg ccc<br>Gln Glu Thr Ser Arg Ile Leu Asp Glu Ser Glu Asp Thr Asp Leu Pro<br>1295 1300 1305      | 4297 |
| tac cca cca ccc cag agg gag gcc aac atc tac atg gtt cct cag aac<br>Tyr Pro Pro Pro Gln Arg Glu Ala Asn Ile Tyr Met Val Pro Gln Asn<br>1310 1315 1320      | 4345 |
| atc aag cca gcg ctc cag cgt acc gcc atc gag atc ctg gca tgg ggc<br>Ile Lys Pro Ala Leu Gln Arg Thr Ala Ile Glu Ile Leu Ala Trp Gly<br>1325 1330 1335 1340 | 4393 |
| ctg ccg aac atg aag agt tac cag ctg gcc aac atc tcc tcc ccc agc<br>Leu Arg Asn Met Lys Ser Tyr Gln Leu Ala Asn Ile Ser Ser Pro Ser<br>1345 1350 1355      | 4441 |
| ctc gtg gta gag tgt ggg ggc cag acg gtg cag tcc tgt gtc atc agg<br>Leu Val Val Glu Cys Gly Gly Gln Thr Val Gln Ser Cys Val Ile Arg<br>1360 1365 1370      | 4489 |
| aac ctc ccg aag aac ccc aac ttt gac atc tgc acc ctc ttc atg gaa<br>Asn Leu Arg Lys Asn Pro Asn Phe Asp Ile Cys Thr Leu Phe Met Glu<br>1375 1380 1385      | 4537 |
| gtg atg ctg ccc agg gag gag ctc tac tgc ccc ccc atc acc gtc aag<br>Val Met Leu Pro Arg Glu Glu Leu Tyr Cys Pro Pro Ile Thr Val Lys<br>1390 1395 1400      | 4585 |
| gtc atc gat aac cgc cag ttt ggc cgc ccg cct gtg gtg ggc cag tgt<br>Val Ile Asp Asn Arg Gln Phe Gly Arg Arg Pro Val Val Gly Gln Cys<br>1405 1410 1415 1420 | 4633 |
| acc atc cgc tcc ctg gag agc ttc ctg tgt gac ccc tac tcc gcg gag<br>Thr Ile Arg Ser Leu Glu Ser Phe Leu Cys Asp Pro Tyr Ser Ala Glu<br>1425 1430 1435      | 4681 |
| agt cca tcc cca cag ggt ggc cca gac gat gtg agc cta ctc agt cct<br>Ser Pro Ser Pro Gln Gly Gly Pro Asp Asp Val Ser Leu Leu Ser Pro<br>1440 1445 1450      | 4729 |

7/68

|   |      |
|---|------|
| ggg gaa gac gtg ctc atc gac att gat gac aag gag ccc ctc atc ccc<br>Gly Glu Asp Val Leu Ile Asp Ile Asp Asp Lys Glu Pro Leu Ile Pro<br>1455 1460 1465      | 4777 |
| atc cag gag gaa gag ttc atc gat tgg tgg agc aaa ttc ttt gcc tcc<br>Ile Gln Glu Glu Glu Phe Ile Asp Trp Trp Ser Lys Phe Phe Ala Ser<br>1470 1475 1480      | 4825 |
| ata ggg gag agg gaa aag tgc ggc tcc tac ctg gag aag gat ttt gac<br>Ile Gly Glu Arg Glu Lys Cys Gly Ser Tyr Leu Glu Lys Asp Phe Asp<br>1485 1490 1495 1500 | 4873 |
| acc ctg aag gtc tat gac aca cag ctg gag aat gtg gag gcc ttt gag<br>Thr Leu Lys Val Tyr Asp Thr Gln Leu Glu Asn Val Glu Ala Phe Glu<br>1505 1510 1515      | 4921 |
| ggc ctg tct gac ttt tgt aac acc ttc aag ctg tac cgg ggc aag acg<br>Gly Leu Ser Asp Phe Cys Asn Thr Phe Lys Leu Tyr Arg Gly Lys Thr<br>1520 1525 1530      | 4969 |
| cag gag gag aca gaa gat cca tct gtg att ggt gaa ttt aag ggc ctc<br>Gln Glu Glu Thr Glu Asp Pro Ser Val Ile Gly Glu Phe Lys Gly Leu<br>1535 1540 1545      | 5017 |
| ttc aaa att tat ccc ctc cca gaa gac cca gcc atc ccc atg ccc cca<br>Phe Lys Ile Tyr Pro Leu Pro Glu Asp Pro Ala Ile Pro Met Pro Pro<br>1550 1555 1560      | 5065 |
| aga cag ttc cac cag ctg gcc gcc cag gga ccc cag gag tgc ttg gtc<br>Arg Gln Phe His Gln Leu Ala Ala Gln Gly Pro Gln Glu Cys Leu Val<br>1565 1570 1575 1580 | 5113 |
| cgt atc tac att gtc cga gca ttt ggc ctg cag ccc aag gac ccc aat<br>Arg Ile Tyr Ile Val Arg Ala Phe Gly Leu Gln Pro Lys Asp Pro Asn<br>1585 1590 1595      | 5161 |
| gga aag tgt gat cct tac atc aag atc tcc ata ggg aag aaa tca gtg<br>Gly Lys Cys Asp Pro Tyr Ile Lys Ile Ser Ile Gly Lys Lys Ser Val<br>1600 1605 1610      | 5209 |
| agt gac cag gat aac tac atc ccc tgc acg ctg gag ccc gta ttt gga<br>Ser Asp Gln Asp Asn Tyr Ile Pro Cys Thr Leu Glu Pro Val Phe Gly<br>1615 1620 1625      | 5257 |
| aag atg ttc gag ctg acc tgc act ctg cct ctg gag aag gac cta aag<br>Lys Met Phe Glu Leu Thr Cys Thr Leu Pro Leu Glu Lys Asp Leu Lys<br>1630 1635 1640      | 5305 |
| atc act ctc tat gac tat gac ctc ctc tcc aag gac gaa aag atc ggt<br>Ile Thr Leu Tyr Asp Tyr Asp Leu Leu Ser Lys Asp Glu Lys Ile Gly<br>1645 1650 1655 1660 | 5353 |
| gag acg gtc gtc gac ctg gag aac agg ctg ctg tcc aag ttt ggg gct<br>Glu Thr Val Val Asp Leu Glu Asn Arg Leu Leu Ser Lys Phe Gly Ala<br>1665 1670 1675      | 5401 |
| cgc tgt gga ctc cca cag acc tac tgt gtc tct gga ccg aac cag tgg<br>Arg Cys Gly Leu Pro Gln Thr Tyr Cys Val Ser Gly Pro Asn Gln Trp<br>1680 1685 1690      | 5449 |
| cgg gac cag ctc cgc ccc tcc cag ctc ctc cac ctc ttc tgc cag cag<br>Arg Asp Gln Leu Arg Pro Ser Gln Leu Leu His Leu Phe Cys Gln Gln<br>1695 1700 1705      | 5497 |

8/68

|   |      |
|---|------|
| cat aga gtc aag gca cct gtg tac cgg aca gac cgt gta atg ttt cag<br>His Arg Val Lys Ala Pro Val Tyr Arg Thr Asp Arg Val Met Phe Gln<br>1710 1715 1720      | 5545 |
| gat aaa gaa tat tcc att gaa gag ata gag gct ggc agg atc cca aac<br>Asp Lys Glu Tyr Ser Ile Glu Glu Ile Glu Ala Gly Arg Ile Pro Asn<br>1725 1730 1735 1740 | 5593 |
| cca cac ctg ggc cca gtg gag gag cgt ctg gct ctg cat gtg ctt cag<br>Pro His Leu Gly Pro Val Glu Glu Arg Leu Ala Leu His Val Leu Gln<br>1745 1750 1755      | 5641 |
| cag cag ggc ctg gtc ccg gag cac gtg gag tca cgg ccc ctc tac agc<br>Gln Gln Gly Leu Val Pro Glu His Val Glu Ser Arg Pro Leu Tyr Ser<br>1760 1765 1770      | 5689 |
| ccc ctg cag cca gac atc gag cag ggg aag ctg cag atg tgg gtc gac<br>Pro Leu Gln Pro Asp Ile Glu Gln Gly Lys Leu Gln Met Trp Val Asp<br>1775 1780 1785      | 5737 |
| cta ttt ccg aag gcc ctg ggg cgg cct gga cct ccc ttc aac atc acc<br>Leu Phe Pro Lys Ala Leu Gly Arg Pro Gly Pro Pro Phe Asn Ile Thr<br>1790 1795 1800      | 5785 |
| cca cgg aga gcc aga agg ttt ttc ctg cgt tgt att atc tgg aat acc<br>Pro Arg Arg Ala Arg Arg Phe Phe Leu Arg Cys Ile Ile Trp Asn Thr<br>1805 1810 1815 1820 | 5833 |
| aga gat gtg atc ctg gat gac ctg agc ctc acg ggg gag aag atg agc<br>Arg Asp Val Ile Leu Asp Asp Leu Ser Leu Thr Gly Glu Lys Met Ser<br>1825 1830 1835      | 5881 |
| gac att tat gtg aaa ggt tgg atg att ggc ttt gaa gaa cac aag caa<br>Asp Ile Tyr Val Lys Gly Trp Met Ile Gly Phe Glu Glu His Lys Gln<br>1840 1845 1850      | 5929 |
| aag aca gac gtg cat tat cgt tcc ctg gga ggt gaa ggc aac ttc aac<br>Lys Thr Asp Val His Tyr Arg Ser Leu Gly Gly Glu Gly Asn Phe Asn<br>1855 1860 1865      | 5977 |
| tgg agg ttc att ttc ccc ttc gac tac ctg cca gct gag caa gtc tgt<br>Trp Arg Phe Ile Phe Pro Phe Asp Tyr Leu Pro Ala Glu Gln Val Cys<br>1870 1875 1880      | 6025 |
| acc att gcc aag aag gat gcc ttc tgg agg ctg gac aag act gag agc<br>Thr Ile Ala Lys Lys Asp Ala Phe Trp Arg Leu Asp Lys Thr Glu Ser<br>1885 1890 1895 1900 | 6073 |
| aaa atc cca gca cga gtg gtg ttc cag atc tgg gac aat gac aag ttc<br>Lys Ile Pro Ala Arg Val Phe Gln Ile Trp Asp Asn Asp Lys Phe<br>1905 1910 1915          | 6121 |
| tcc ttt gat gat ttt ctg ggc tcc ctg cag ctc gat ctc aac cgc atg<br>Ser Phe Asp Asp Phe Leu Gly Ser Leu Gln Leu Asp Leu Asn Arg Met<br>1920 1925 1930      | 6169 |
| ccc aag cca gcc aag aca gcc aag aag tgc tcc ttg gac cag ctg gat<br>Pro Lys Pro Ala Lys Thr Ala Lys Lys Cys Ser Leu Asp Gln Leu Asp<br>1935 1940 1945      | 6217 |
| gat gct ttc cac cca gaa tgg ttt gtg tcc ctt ttt gag cag aaa aca<br>Asp Ala Phe His Pro Glu Trp Phe Val Ser Leu Phe Glu Gln Lys Thr<br>1950 1955 1960      | 6265 |
| gtg aag ggc tgg tgg ccc tgt gta gca gaa gag ggt gag aag aaa ata<br>Val Lys Gly Trp Trp Pro Cys Val Ala Glu Glu Gly Glu Lys Lys Ile<br>1965 1970 1975 1980 | 6313 |

9/68

ctg gcg ggc aag ctg gaa atg acc ttg gag att gta gca gag agt gag 6361  
 Leu Ala Gly Lys Leu Glu Met Thr Leu Glu Ile Val Ala Glu Ser Glu  
 1985 1990 1995  
 cat gag gag cgg cct gct ggc cag ggc cgg gat gag ccc aac atg aac 6409  
 His Glu Glu Arg Pro Ala Gly Gln Gly Arg Asp Glu Pro Asn Met Asn  
 2000 2005 2010  
 cct aag ctt gag gac cca agg cgc ccc gac acc tcc ttc ctg tgg ttt 6457  
 Pro Lys Leu Glu Asp Pro Arg Arg Pro Asp Thr Ser Phe Leu Trp Phe  
 2015 2020 2025  
 acc tcc cca tac aag acc atg aag ttc atc ctg tgg cgg cgt ttc cgg 6505  
 Thr Ser Pro Tyr Lys Thr Met Lys Phe Ile Leu Trp Arg Arg Phe Arg  
 2030 2035 2040  
 tgg gcc atc atc ctc ttc atc atc ctc ttc atc ctg ctg ctg ttc ctg 6553  
 Trp Ala Ile Ile Leu Phe Ile Ile Leu Phe Ile Leu Leu Leu Phe Leu  
 2045 2050 2055 2060  
 gcc atc ttc atc tac gcc ttc cgg aac tat gct gcc atg aag ctg gtg 6601  
 Ala Ile Phe Ile Tyr Ala Phe Pro Asn Tyr Ala Ala Met Lys Leu Val  
 2065 2070 2075  
 aag ccc ttc agc tgaggactct cctgccctgt agaagggggcc gtgggggtccc 6653  
 Lys Pro Phe Ser  
 2080  
 ctccagcatg ggactggcct gcctcctccg cccagctcgg cgagctcctc cagacctcct 6713  
 aggcctgatt gtcctgccag ggtgggcaga cagacagatg gaccggccca cactcccaga 6773  
 gttgctaaca tggagctctg agatcacccc acttccatca tttccttctc ccccaaccca 6833  
 acgctttttt ggatcagctc agacatattt cagtataaaa cagttggaac cacaacccca 6893  
 aaaaaaaaaa aaaaaaaaaa 6911

<210> 2  
 <211> 2080  
 <212> PRT  
 <213> Homo sapiens

<400> 2  
 Met Leu Arg Val Phe Ile Leu Tyr Ala Glu Asn Val His Thr Pro Asp  
 1 5 10 15  
 Thr Asp Ile Ser Asp Ala Tyr Cys Ser Ala Val Phe Ala Gly Val Lys  
 20 25 30  
 Lys Arg Thr Lys Val Ile Lys Asn Ser Val Asn Pro Val Trp Asn Glu  
 35 40 45  
 Gly Phe Glu Trp Asp Leu Lys Gly Ile Pro Leu Asp Gln Gly Ser Glu  
 50 55 60  
 Leu His Val Val Val Lys Asp His Glu Thr Met Gly Arg Asn Arg Phe  
 65 70 75 80  
 Leu Gly Glu Ala Lys Val Pro Leu Arg Glu Val Leu Ala Thr Pro Ser  
 85 90 95  
 Leu Ser Ala Ser Phe Asn Ala Pro Leu Leu Asp Thr Lys Lys Gln Pro  
 100 105 110  
 Thr Gly Ala Ser Leu Val Leu Gln Val Ser Tyr Thr Pro Leu Pro Gly  
 115 120 125  
 Ala Val Pro Leu Phe Pro Pro Thr Pro Leu Glu Pro Ser Pro Thr  
 130 135 140  
 Leu Pro Asp Leu Asp Val Val Ala Asp Thr Gly Gly Glu Glu Asp Thr  
 145 150 155 160  
 Glu Asp Gln Gly Leu Thr Gly Asp Glu Ala Glu Pro Phe Leu Asp Gln  
 165 170 175  
 Ser Gly Gly Pro Gly Ala Pro Thr Thr Pro Arg Lys Leu Pro Ser Arg  
 180 185 190  
 Pro Pro Pro His Tyr Pro Gly Ile Lys Arg Lys Arg Ser Ala Pro Thr  
 195 200 205



10/68

Ser Arg Lys Leu Leu Ser Asp Lys Pro Gln Asp Phe Gln Ile Arg Val  
 210 215 220  
 Gln Val Ile Glu Gly Arg Gln Leu Pro Gly Val Asn Ile Lys Pro Val  
 225 230 235 240  
 Val Lys Val Thr Ala Ala Gly Gln Thr Lys Arg Thr Arg Ile His Lys  
 245 250 255  
 Gly Asn Ser Pro Leu Phe Asn Glu Thr Leu Phe Phe Asn Leu Phe Asp  
 260 265 270  
 Ser Pro Gly Glu Leu Phe Asp Glu Pro Ile Phe Ile Thr Val Val Asp  
 275 280 285  
 Ser Arg Ser Leu Arg Thr Asp Ala Leu Leu Gly Glu Phe Arg Met Asp  
 290 295 300  
 Val Gly Thr Ile Tyr Arg Glu Pro Arg His Ala Tyr Leu Arg Lys Trp  
 305 310 315 320  
 Leu Leu Leu Ser Asp Pro Asp Asp Phe Ser Ala Gly Ala Arg Gly Tyr  
 325 330 335  
 Leu Lys Thr Ser Leu Cys Val Leu Gly Pro Gly Asp Glu Ala Pro Leu  
 340 345 350  
 Glu Arg Lys Asp Pro Ser Glu Asp Lys Glu Asp Ile Glu Ser Asn Leu  
 355 360 365  
 Leu Arg Pro Thr Gly Val Ala Leu Arg Gly Ala His Phe Cys Leu Lys  
 370 375 380  
 Val Phe Arg Ala Glu Asp Leu Pro Gln Met Asp Asp Ala Val Met Asp  
 385 390 395 400  
 Asn Val Lys Gln Ile Phe Gly Phe Glu Ser Asn Lys Lys Asn Leu Val  
 405 410 415  
 Asp Pro Phe Val Glu Val Ser Phe Ala Gly Lys Met Leu Cys Ser Lys  
 420 425 430  
 Ile Leu Glu Lys Thr Ala Asn Pro Gln Trp Asn Gln Asn Ile Thr Leu  
 435 440 445  
 Pro Ala Met Phe Pro Ser Met Cys Glu Lys Met Arg Ile Arg Ile Ile  
 450 455 460  
 Asp Trp Asp Arg Leu Thr His Asn Asp Ile Val Ala Thr Thr Tyr Leu  
 465 470 475 480  
 Ser Met Ser Lys Ile Ser Ala Pro Gly Gly Glu Ile Glu Glu Glu Pro  
 485 490 495  
 Ala Gly Ala Val Lys Pro Ser Lys Ala Ser Asp Leu Asp Asp Tyr Leu  
 500 505 510  
 Gly Phe Leu Pro Thr Phe Gly Pro Cys Tyr Ile Asn Leu Tyr Gly Ser  
 515 520 525  
 Pro Arg Glu Phe Thr Gly Phe Pro Asp Pro Tyr Thr Glu Leu Asn Thr  
 530 535 540  
 Gly Lys Gly Glu Gly Val Ala Tyr Arg Gly Arg Leu Leu Leu Ser Leu  
 545 550 555 560  
 Glu Thr Lys Leu Val Glu His Ser Glu Gln Lys Val Glu Asp Leu Pro  
 565 570 575  
 Ala Asp Asp Ile Leu Arg Val Glu Lys Tyr Leu Arg Arg Arg Lys Tyr  
 580 585 590  
 Ser Leu Phe Ala Ala Phe Tyr Ser Ala Thr Met Leu Gln Asp Val Asp  
 595 600 605  
 Asp Ala Ile Gln Phe Glu Val Ser Ile Gly Asn Tyr Gly Asn Lys Phe  
 610 615 620  
 Asp Met Thr Cys Leu Pro Leu Ala Ser Thr Thr Gln Tyr Ser Arg Ala  
 625 630 635 640  
 Val Phe Asp Gly Cys His Tyr Tyr Tyr Leu Pro Trp Gly Asn Val Lys  
 645 650 655  
 Pro Val Val Val Leu Ser Ser Tyr Trp Glu Asp Ile Ser His Arg Ile  
 660 665 670  
 Glu Thr Gln Asn Gln Leu Leu Gly Ile Ala Asp Arg Leu Glu Ala Gly  
 675 680 685  
 Leu Glu Gln Val His Leu Ala Leu Lys Ala Gln Cys Ser Thr Glu Asp  
 690 695 700  
 Val Asp Ser Leu Val Ala Gln Leu Thr Asp Glu Leu Ile Ala Gly Cys  
 705 710 715 720  
 Ser Gln Pro Leu Gly Asp Ile His Glu Thr Pro Ser Ala Thr His Leu  
 725 730 735

11/68

|      |      |      |      |     |      |      |      |      |      |      |      |      |      |      |      |  |  |
|------|------|------|------|-----|------|------|------|------|------|------|------|------|------|------|------|--|--|
| Asp  | Gln  | Tyr  | Leu  | Tyr | Gln  | Leu  | Arg  | Thr  | His  | His  | Leu  | Ser  | Gln  | Ile  | Thr  |  |  |
|      |      |      | 740  |     |      |      |      | 745  |      |      |      |      | 750  |      |      |  |  |
| Glu  | Ala  | Ala  | Leu  | Ala | Leu  | Lys  | Leu  | Gly  | His  | Ser  | Glu  | Leu  | Pro  | Ala  | Ala  |  |  |
|      |      | 755  |      |     |      |      | 760  |      |      |      |      | 765  |      |      |      |  |  |
| Leu  | Glu  | Gln  | Ala  | Glu | Asp  | Trp  | Leu  | Leu  | Arg  | Leu  | Arg  | Ala  | Leu  | Ala  | Glu  |  |  |
|      | 770  |      |      |     |      | 775  |      |      |      |      | 780  |      |      |      |      |  |  |
| Glu  | Pro  | Gln  | Asn  | Ser | Leu  | Pro  | Asp  | Ile  | Val  | Ile  | Trp  | Met  | Leu  | Gln  | Gly  |  |  |
| 785  |      |      |      |     | 790  |      |      |      |      | 795  |      |      |      |      | 800  |  |  |
| Asp  | Lys  | Arg  | Val  | Ala | Tyr  | Gln  | Arg  | Val  | Pro  | Ala  | His  | Gln  | Val  | Leu  | Phe  |  |  |
|      |      |      | 805  |     |      |      |      |      | 810  |      |      |      |      | 815  |      |  |  |
| Ser  | Arg  | Arg  | Gly  | Ala | Asn  | Tyr  | Cys  | Gly  | Lys  | Asn  | Cys  | Gly  | Lys  | Leu  | Gln  |  |  |
|      |      |      | 820  |     |      |      |      | 825  |      |      |      |      | 830  |      |      |  |  |
| Thr  | Ile  | Phe  | Leu  | Lys | Tyr  | Pro  | Met  | Glu  | Lys  | Val  | Pro  | Gly  | Ala  | Arg  | Met  |  |  |
|      | 835  |      |      |     |      |      | 840  |      |      |      |      | 845  |      |      |      |  |  |
| Pro  | Val  | Gln  | Ile  | Arg | Val  | Lys  | Leu  | Trp  | Phe  | Gly  | Leu  | Ser  | Val  | Asp  | Glu  |  |  |
|      | 850  |      |      |     |      | 855  |      |      |      |      | 860  |      |      |      |      |  |  |
| Lys  | Glu  | Phe  | Asn  | Gln | Phe  | Ala  | Glu  | Gly  | Lys  | Leu  | Ser  | Val  | Phe  | Ala  | Glu  |  |  |
| 865  |      |      |      |     | 870  |      |      |      |      | 875  |      |      |      |      | 880  |  |  |
| Thr  | Tyr  | Glu  | Asn  | Glu | Thr  | Lys  | Leu  | Ala  | Leu  | Val  | Gly  | Asn  | Trp  | Gly  | Thr  |  |  |
|      |      |      | 885  |     |      |      |      |      | 890  |      |      |      |      | 895  |      |  |  |
| Thr  | Gly  | Leu  | Thr  | Tyr | Pro  | Lys  | Phe  | Ser  | Asp  | Val  | Thr  | Gly  | Lys  | Ile  | Lys  |  |  |
|      |      | 900  |      |     |      |      |      | 905  |      |      |      |      | 910  |      |      |  |  |
| Leu  | Pro  | Lys  | Asp  | Ser | Phe  | Arg  | Pro  | Ser  | Ala  | Gly  | Trp  | Thr  | Trp  | Ala  | Gly  |  |  |
|      |      | 915  |      |     |      |      | 920  |      |      |      |      | 925  |      |      |      |  |  |
| Asp  | Trp  | Phe  | Val  | Cys | Pro  | Glu  | Lys  | Thr  | Leu  | Leu  | His  | Asp  | Met  | Asp  | Ala  |  |  |
|      | 930  |      |      |     |      | 935  |      |      |      |      | 940  |      |      |      |      |  |  |
| Gly  | His  | Leu  | Ser  | Phe | Val  | Glu  | Glu  | Val  | Phe  | Glu  | Asn  | Gln  | Thr  | Arg  | Leu  |  |  |
| 945  |      |      |      |     | 950  |      |      |      |      | 955  |      |      |      |      | 960  |  |  |
| Pro  | Gly  | Gly  | Gln  | Trp | Ile  | Tyr  | Met  | Ser  | Asp  | Asn  | Tyr  | Thr  | Asp  | Val  | Asn  |  |  |
|      |      |      | 965  |     |      |      |      |      | 970  |      |      |      |      | 975  |      |  |  |
| Gly  | Glu  | Lys  | Val  | Leu | Pro  | Lys  | Asp  | Asp  | Ile  | Glu  | Cys  | Pro  | Leu  | Gly  | Trp  |  |  |
|      |      | 980  |      |     |      |      | 985  |      |      |      |      |      | 990  |      |      |  |  |
| Lys  | Trp  | Glu  | Asp  | Glu | Glu  | Trp  | Ser  | Thr  | Asp  | Leu  | Asn  | Arg  | Ala  | Val  | Asp  |  |  |
|      |      | 995  |      |     |      |      | 1000 |      |      |      |      | 1005 |      |      |      |  |  |
| Glu  | Gln  | Gly  | Trp  | Glu | Tyr  | Ser  | Ile  | Thr  | Ile  | Pro  | Pro  | Glu  | Arg  | Lys  | Pro  |  |  |
|      | 1010 |      |      |     |      | 1015 |      |      |      | 1020 |      |      |      |      |      |  |  |
| Lys  | His  | Trp  | Val  | Pro | Ala  | Glu  | Lys  | Met  | Tyr  | Tyr  | Thr  | His  | Arg  | Arg  | Arg  |  |  |
| 1025 |      |      |      |     | 1030 |      |      |      | 1035 |      |      |      |      |      | 1040 |  |  |
| Arg  | Trp  | Val  | Arg  | Leu | Arg  | Arg  | Arg  | Asp  | Leu  | Ser  | Gln  | Met  | Glu  | Ala  | Leu  |  |  |
|      |      |      | 1045 |     |      |      |      |      | 1050 |      |      |      |      | 1055 |      |  |  |
| Lys  | Arg  | His  | Arg  | Gln | Ala  | Glu  | Ala  | Glu  | Gly  | Glu  | Gly  | Trp  | Glu  | Tyr  | Ala  |  |  |
|      |      | 1060 |      |     |      |      |      | 1065 |      |      |      |      | 1070 |      |      |  |  |
| Ser  | Leu  | Phe  | Gly  | Trp | Lys  | Phe  | His  | Leu  | Glu  | Tyr  | Arg  | Lys  | Thr  | Asp  | Ala  |  |  |
|      |      | 1075 |      |     |      |      | 1080 |      |      |      |      | 1085 |      |      |      |  |  |
| Phe  | Arg  | Arg  | Arg  | Arg | Trp  | Arg  | Arg  | Arg  | Met  | Glu  | Pro  | Leu  | Glu  | Lys  | Thr  |  |  |
|      | 1090 |      |      |     |      | 1095 |      |      |      | 1100 |      |      |      |      |      |  |  |
| Gly  | Pro  | Ala  | Ala  | Val | Phe  | Ala  | Leu  | Glu  | Gly  | Ala  | Leu  | Gly  | Gly  | Val  | Met  |  |  |
| 1105 |      |      |      |     | 1110 |      |      |      |      | 1115 |      |      |      |      | 1120 |  |  |
| Asp  | Asp  | Lys  | Ser  | Glu | Asp  | Ser  | Met  | Ser  | Val  | Ser  | Thr  | Leu  | Ser  | Phe  | Gly  |  |  |
|      |      |      | 1125 |     |      |      |      |      | 1130 |      |      |      |      | 1135 |      |  |  |
| Val  | Asn  | Arg  | Pro  | Thr | Ile  | Ser  | Cys  | Ile  | Phe  | Asp  | Tyr  | Gly  | Asn  | Arg  | Tyr  |  |  |
|      |      | 1140 |      |     |      |      |      | 1145 |      |      |      |      | 1150 |      |      |  |  |
| His  | Leu  | Arg  | Cys  | Tyr | Met  | Tyr  | Gln  | Ala  | Arg  | Asp  | Leu  | Ala  | Ala  | Met  | Asp  |  |  |
|      |      | 1155 |      |     |      |      | 1160 |      |      |      |      | 1165 |      |      |      |  |  |
| Lys  | Asp  | Ser  | Phe  | Ser | Asp  | Pro  | Tyr  | Ala  | Ile  | Val  | Ser  | Phe  | Leu  | His  | Gln  |  |  |
|      | 1170 |      |      |     |      | 1175 |      |      |      |      | 1180 |      |      |      |      |  |  |
| Ser  | Gln  | Lys  | Thr  | Val | Val  | Val  | Lys  | Asn  | Thr  | Leu  | Asn  | Pro  | Thr  | Trp  | Asp  |  |  |
| 1185 |      |      |      |     | 1190 |      |      |      |      | 1195 |      |      |      |      | 1200 |  |  |
| Gln  | Thr  | Leu  | Ile  | Phe | Tyr  | Glu  | Ile  | Glu  | Ile  | Phe  | Gly  | Glu  | Pro  | Ala  | Thr  |  |  |
|      |      |      | 1205 |     |      |      |      |      | 1210 |      |      |      |      | 1215 |      |  |  |
| Val  | Ala  | Glu  | Gln  | Pro | Pro  | Ser  | Ile  | Val  | Val  | Glu  | Leu  | Tyr  | Asp  | His  | Asp  |  |  |
|      |      | 1220 |      |     |      |      |      | 1225 |      |      |      |      | 1230 |      |      |  |  |
| Thr  | Tyr  | Gly  | Ala  | Asp | Glu  | Phe  | Met  | Gly  | Arg  | Cys  | Ile  | Cys  | Gln  | Pro  | Ser  |  |  |
|      |      | 1235 |      |     |      |      | 1240 |      |      |      |      | 1245 |      |      |      |  |  |
| Leu  | Glu  | Arg  | Met  | Pro | Arg  | Leu  | Ala  | Trp  | Phe  | Pro  | Leu  | Thr  | Arg  | Gly  | Ser  |  |  |
|      | 1250 |      |      |     |      | 1255 |      |      |      |      | 1260 |      |      |      |      |  |  |

12/68

Gln Pro Ser Gly Glu Leu Leu Ala Ser Phe Glu Leu Ile Gln Arg Glu  
 1265 1270 1275 1280  
 Lys Pro Ala Ile His His Ile Pro Gly Phe Glu Val Gln Glu Thr Ser  
 1285 1290 1295  
 Arg Ile Leu Asp Glu Ser Glu Asp Thr Asp Leu Pro Tyr Pro Pro Pro  
 1300 1305 1310  
 Gln Arg Glu Ala Asn Ile Tyr Met Val Pro Gln Asn Ile Lys Pro Ala  
 1315 1320 1325  
 Leu Gln Arg Thr Ala Ile Glu Ile Leu Ala Trp Gly Leu Arg Asn Met  
 1330 1335 1340  
 Lys Ser Tyr Gln Leu Ala Asn Ile Ser Ser Pro Ser Leu Val Val Glu  
 1345 1350 1355 1360  
 Cys Gly Gly Gln Thr Val Gln Ser Cys Val Ile Arg Asn Leu Arg Lys  
 1365 1370 1375  
 Asn Pro Asn Phe Asp Ile Cys Thr Leu Phe Met Glu Val Met Leu Pro  
 1380 1385 1390  
 Arg Glu Glu Leu Tyr Cys Pro Pro Ile Thr Val Lys Val Ile Asp Asn  
 1395 1400 1405  
 Arg Gln Phe Gly Arg Arg Pro Val Val Gly Gln Cys Thr Ile Arg Ser  
 1410 1415 1420  
 Leu Glu Ser Phe Leu Cys Asp Pro Tyr Ser Ala Glu Ser Pro Ser Pro  
 1425 1430 1435 1440  
 Gln Gly Gly Pro Asp Asp Val Ser Leu Leu Ser Pro Gly Glu Asp Val  
 1445 1450 1455  
 Leu Ile Asp Ile Asp Asp Lys Glu Pro Leu Ile Pro Ile Gln Glu Glu  
 1460 1465 1470  
 Glu Phe Ile Asp Trp Trp Ser Lys Phe Phe Ala Ser Ile Gly Glu Arg  
 1475 1480 1485  
 Glu Lys Cys Gly Ser Tyr Leu Glu Lys Asp Phe Asp Thr Leu Lys Val  
 1490 1495 1500  
 Tyr Asp Thr Gln Leu Glu Asn Val Glu Ala Phe Glu Gly Leu Ser Asp  
 1505 1510 1515 1520  
 Phe Cys Asn Thr Phe Lys Leu Tyr Arg Gly Lys Thr Gln Glu Glu Thr  
 1525 1530 1535  
 Glu Asp Pro Ser Val Ile Gly Glu Phe Lys Gly Leu Phe Lys Ile Tyr  
 1540 1545 1550  
 Pro Leu Pro Glu Asp Pro Ala Ile Pro Met Pro Pro Arg Gln Phe His  
 1555 1560 1565  
 Gln Leu Ala Ala Gln Gly Pro Gln Glu Cys Leu Val Arg Ile Tyr Ile  
 1570 1575 1580  
 Val Arg Ala Phe Gly Leu Gln Pro Lys Asp Pro Asn Gly Lys Cys Asp  
 1585 1590 1595 1600  
 Pro Tyr Ile Lys Ile Ser Ile Gly Lys Lys Ser Val Ser Asp Gln Asp  
 1605 1610 1615  
 Asn Tyr Ile Pro Cys Thr Leu Glu Pro Val Phe Gly Lys Met Phe Glu  
 1620 1625 1630  
 Leu Thr Cys Thr Leu Pro Leu Glu Lys Asp Leu Lys Ile Thr Leu Tyr  
 1635 1640 1645  
 Asp Tyr Asp Leu Leu Ser Lys Asp Glu Lys Ile Gly Glu Thr Val Val  
 1650 1655 1660  
 Asp Leu Glu Asn Arg Leu Leu Ser Lys Phe Gly Ala Arg Cys Gly Leu  
 1665 1670 1675 1680  
 Pro Gln Thr Tyr Cys Val Ser Gly Pro Asn Gln Trp Arg Asp Gln Leu  
 1685 1690 1695  
 Arg Pro Ser Gln Leu Leu His Leu Phe Cys Gln Gln His Arg Val Lys  
 1700 1705 1710  
 Ala Pro Val Tyr Arg Thr Asp Arg Val Met Phe Gln Asp Lys Glu Tyr  
 1715 1720 1725  
 Ser Ile Glu Glu Ile Glu Ala Gly Arg Ile Pro Asn Pro His Leu Gly  
 1730 1735 1740  
 Pro Val Glu Glu Arg Leu Ala Leu His Val Leu Gln Gln Gln Gly Leu  
 1745 1750 1755 1760  
 Val Pro Glu His Val Glu Ser Arg Pro Leu Tyr Ser Pro Leu Gln Pro  
 1765 1770 1775  
 Asp Ile Glu Gln Gly Lys Leu Gln Met Trp Val Asp Leu Phe Pro Lys  
 1780 1785 1790

13/68

Ala Leu Gly Arg Pro Gly Pro Pro Phe Asn Ile Thr Pro Arg Arg Ala  
 1795 1800 1805  
 Arg Arg Phe Phe Leu Arg Cys Ile Ile Trp Asn Thr Arg Asp Val Ile  
 1810 1815 1820  
 Leu Asp Asp Leu Ser Leu Thr Gly Glu Lys Met Ser Asp Ile Tyr Val  
 1825 1830 1835 1840  
 Lys Gly Trp Met Ile Gly Phe Glu Glu His Lys Gln Lys Thr Asp Val  
 1845 1850 1855  
 His Tyr Arg Ser Leu Gly Gly Glu Gly Asn Phe Asn Trp Arg Phe Ile  
 1860 1865 1870  
 Phe Pro Phe Asp Tyr Leu Pro Ala Glu Gln Val Cys Thr Ile Ala Lys  
 1875 1880 1885  
 Lys Asp Ala Phe Trp Arg Leu Asp Lys Thr Glu Ser Lys Ile Pro Ala  
 1890 1895 1900  
 Arg Val Val Phe Gln Ile Trp Asp Asn Asp Lys Phe Ser Phe Asp Asp  
 1905 1910 1915 1920  
 Phe Leu Gly Ser Leu Gln Leu Asp Leu Asn Arg Met Pro Lys Pro Ala  
 1925 1930 1935  
 Lys Thr Ala Lys Lys Cys Ser Leu Asp Gln Leu Asp Asp Ala Phe His  
 1940 1945 1950  
 Pro Glu Trp Phe Val Ser Leu Phe Glu Gln Lys Thr Val Lys Gly Trp  
 1955 1960 1965  
 Trp Pro Cys Val Ala Glu Glu Gly Glu Lys Lys Ile Leu Ala Gly Lys  
 1970 1975 1980  
 Leu Glu Met Thr Leu Glu Ile Val Ala Glu Ser Glu His Glu Glu Arg  
 1985 1990 1995 2000  
 Pro Ala Gly Gln Gly Arg Asp Glu Pro Asn Met Asn Pro Lys Leu Glu  
 2005 2010 2015  
 Asp Pro Arg Arg Pro Asp Thr Ser Phe Leu Trp Phe Thr Ser Pro Tyr  
 2020 2025 2030  
 Lys Thr Met Lys Phe Ile Leu Trp Arg Arg Phe Arg Trp Ala Ile Ile  
 2035 2040 2045  
 Leu Phe Ile Ile Leu Phe Ile Leu Leu Leu Phe Leu Ala Ile Phe Ile  
 2050 2055 2060  
 Tyr Ala Phe Pro Asn Tyr Ala Ala Met Lys Leu Val Lys Pro Phe Ser  
 2065 2070 2075 2080

<210> 3  
 <211> 5915  
 <212> DNA  
 <213> Homo sapiens

<400> 3  
 tcgaccgccc agccaggtgc aaaatgccgt gtcattggga gactccgcag ccggagcatt 60  
 agattacagc tcgacggagc tcgggaagg ggcgggggt ggaagatgag cagaagcccc 120  
 tgttctcgga acgcccgtg acaagcggg tgagcgcagg cggggcgggg acccagccta 180  
 gcccactgga gcagccggg gtggcccgtt cccctttaag agcaactgct ctaagccagg 240  
 agccagagat tcgagccggc ctgcgccagc cagccctctc cagcagaggg acccacaagc 300  
 ggcgcctcgg cctcccgac ctttccgagc cctctttgcg cctgggcgc acggggccct 360  
 acacgcgcca agcatgctga gggtcttcat cctctatgcc gagaacgtcc acacaccga 420  
 caccgacatc agcgatgcct actgctccgc ggtgtttgca ggggtgaaga agagaaccaa 480  
 agtcatcaag aacagcgtga accctgtatg gaatgaggga tttgaatggg acctcaaggg 540  
 catccccctg gaccagggt ctgagcttca tgtgggtggc aaagaccatg agacgatggg 600  
 gaggaacagg ttcttgggg aagccaagg cccactccga gaggtcctcg ccaccctag 660  
 tctgtccgcc agcttcaatg cccccctgct ggacaccaag aagcagccca caggggcctc 720  
 gctgggtcctg caggtgtcct acacaccgct gcctggagct gtgcccctgt tcccgcctcc 780  
 tactcctctg gagccctccc cgactctgcc tgacctggat gtatggcag acacaggagg 840  
 agaggaagac acagaggacc agggactcac tggagatgag gcggagccat tcttggatca 900  
 aagcggaggc ccgggggctc ccaccacccc aaggaaacta ccttcacgtc ctcggcccca 960  
 ctaccccggg atcaaaaagaa agcgaagtgc gcctacatct agaaagctgc tgtcagacaa 1020  
 accgcaggat ttccagatca gggtccagg gatcgagggg cgccagctgc cgggggtgaa 1080  
 catcaagcct gtggtcaagg ttaccgctgc agggcagacc aagcggacgc ggatccacaa 1140  
 gggaaacagc ccactcttca atgagactct tttcttcaac ttgtttgact ctctgggga 1200  
 gctgtttgat gagccatct ttatcacggt ggtagactct cgttctctca ggacagatgc 1260  
 tctcctcggg gatttcggga tggacgtggg caccatttac agagagcccc ggcacgccta 1320  
 tctcaggaag tggctgctgc tctcagaccc tgatgacttc tctgctgggg ccagaggcta 1380  
 cctgaaaaca agcctttgtg tgctggggcc tggggacgaa gcgcctctgg agagaaaaga 1440

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| ccccctctgaa | gacaaggagg  | acattgaaag  | caacctgctc  | cggcccacag  | gcgtagccct  | 1500 |
| gcgaggagcc  | cacttctgcc  | tgaaggctct  | ccgggcccag  | gacttgccgc  | agatggacga  | 1560 |
| tgccgtgatg  | gacaacgtga  | aacagatctt  | tggcttcgag  | agtaacaaga  | agaacttggt  | 1620 |
| ggaccccttt  | gtggaggtca  | gctttgcggg  | gaaaatgctg  | tgcagcaaga  | tcttgagaa   | 1680 |
| gacggccaac  | cctcagtggg  | accagaacat  | cacactgcct  | gccatgtttc  | cctccatgtg  | 1740 |
| cgaaaaaatg  | aggattcgta  | tcatagactg  | ggaccgcctg  | actcacaatg  | acatcgtggc  | 1800 |
| taccacctac  | ctgagtatgt  | cgaaaaatctc | tgccccctgga | ggagaaatag  | aagaggagcc  | 1860 |
| tgcaggtgct  | gtcaagcctt  | cgaaagcctc  | agacttggtg  | gactacctgg  | gcttccctccc | 1920 |
| cacttttggg  | ccctgctaca  | tcaacctcta  | tggcagtcct  | agagagttca  | caggcttccc  | 1980 |
| agacccttac  | acagagctca  | acacaggcaa  | gggggaaggt  | gtggcttatc  | gtggccggct  | 2040 |
| tctgctctcc  | ctggagacca  | agctggtgga  | gcacagtga   | cagaaggtgg  | aggaccttcc  | 2100 |
| tgccgatgac  | atcctccggg  | tggagaagta  | ccttaggagg  | cgcaagtact  | ccctgtttgc  | 2160 |
| ggccttctac  | tcagccacca  | tgtgcagga   | tgtggatgat  | gccatccagt  | ttgaggtcag  | 2220 |
| catcggaac   | tacgggaaca  | agttcgacat  | gacctgcctg  | ccgctggcct  | ccaccactca  | 2280 |
| gtacagccgt  | gcagtctttg  | acgggtgcca  | ctactactac  | ctaccctggg  | gtaacgtgaa  | 2340 |
| acctgtgggt  | gtgctgtcat  | cctactggga  | ggacatcagc  | catagaatcg  | agactcagaa  | 2400 |
| ccagctgctt  | gggattgctg  | accggctgga  | agctggcctg  | gagcaggtcc  | acctggccct  | 2460 |
| gaaggcgag   | tgtctccacg  | aggacgtgga  | ctcgctggtg  | gctcagctga  | cggatgagct  | 2520 |
| catcgaggg   | tgcagccagc  | ctctgggtga  | catccatgag  | acaccctctg  | ccaccacct   | 2580 |
| ggaccagtac  | ctgtaccagc  | tgcgcaccca  | tcacctgagc  | caaactcactg | aggctgcct   | 2640 |
| ggccctgagc  | ctcgccaca   | gtgagctccc  | tgcagctctg  | gagcaggcgg  | aggactggct  | 2700 |
| cctgcgtctg  | cgtgccctgg  | cagaggagcc  | ccagaacagc  | ctggccgaca  | tcgtcatctg  | 2760 |
| gatgctgcag  | ggagacaagc  | gtgtggcata  | ccagcgggtg  | cccgccacc   | aagtccctct  | 2820 |
| ctcccggcgg  | ggtgccaaact | actgtggcaa  | gaattgtggg  | aagctacaga  | caatctttct  | 2880 |
| gaaatatccg  | atggagaagg  | tgcctggcgc  | ccggatgcc   | gtgcagatac  | gggtcaagct  | 2940 |
| gtgggttggg  | ctctctgtgg  | atgagaagga  | gttcaaccag  | tttgcctgag  | ggaagctgtc  | 3000 |
| tgtctttgct  | gaaacctatg  | agaacgagac  | taagtggcc   | cttggtggga  | actggggcac  | 3060 |
| aacgggcctc  | acctaccca   | agttttctga  | cgtcacgggc  | aagatcaagc  | taccaagga   | 3120 |
| cagcttccgc  | ccctcgcccg  | gctggacctg  | ggctggagat  | tggttcgtgt  | gtccggagaa  | 3180 |
| gactctgctc  | catgacatgg  | acgcccgtca  | cctgagcttc  | gtggaagagg  | tggttgagaa  | 3240 |
| ccagaccgg   | cttcccggag  | gccagtggat  | ctacatgagt  | gacaactaca  | ccgatgtgaa  | 3300 |
| cggggagaa   | gtgcttccca  | aggatgacat  | tgagtgccca  | ctgggctgga  | agtgggaaga  | 3360 |
| tgaggaatgg  | tccacagacc  | tcaaccgggc  | tgtcgtatgag | caaggctggg  | agtatagcat  | 3420 |
| caccatcccc  | ccggagcggg  | agccgaagca  | ctgggtccct  | gctgagaaga  | tgtactacac  | 3480 |
| acaccgacgg  | cggcgctggg  | tgcgcctgcg  | caggagggat  | ctcagccaaa  | tggaaagcact | 3540 |
| gaaaaggcac  | aggcaggcgg  | aggcggaggg  | cgagggctgg  | gagtagcctc  | ctcttttttg  | 3600 |
| ctggaagtcc  | cacctcgagt  | accgcaagac  | agatgccttc  | cgccgcgcgc  | gctggcgccg  | 3660 |
| tcgcatggag  | ccactggaga  | agacggggcc  | tgcagctgtg  | tttgcccttg  | agggggccct  | 3720 |
| ggcgggcgtg  | atggatgaca  | agagtgaaga  | ttccatgtcc  | gtctccacct  | tgagcttcgg  | 3780 |
| tgtgaacaga  | cccacgattt  | cctgcatatt  | cgactatggg  | aaccgctacc  | atctacgctg  | 3840 |
| ctacatgtac  | caggccccgg  | acctggctgc  | gatggacaag  | gactcttttt  | ctgatcccta  | 3900 |
| tgccatcgct  | tccttccctg  | accagagcca  | gaagacgggt  | gtggtgaaga  | acacccttaa  | 3960 |
| ccccacctgg  | gaccagacgc  | tcattcttcta | cgagatcgag  | atctttggcg  | agccggccac  | 4020 |
| agttgctgag  | caaccgcccc  | gcattgtggt  | ggagctgtac  | gaccatgaca  | cttatggtgc  | 4080 |
| agacgagttt  | atgggtcgct  | gcattctgtca | accgagctct  | gaacggatgc  | cacggctggc  | 4140 |
| ctggttccca  | ctgacgaggg  | gcagccagcc  | gtcgggggag  | ctgctggcct  | cttttgagct  | 4200 |
| catccagaga  | gagaagccgg  | ccatccacca  | tattcctggt  | tttgaggtgc  | aggagacatc  | 4260 |
| aaggatcctg  | gatgagtctg  | aggacacaga  | cctgccttac  | ccaccacccc  | agagggagcc  | 4320 |
| caacatctac  | atggttcctc  | agaacatcaa  | gccagcgctc  | cagcgtaccg  | ccatcgagat  | 4380 |
| cctggcatgg  | ggcctgcgga  | acatgaagag  | ttaccagctg  | gccaacatct  | cctccccag   | 4440 |
| cctcggtgta  | gagtgtgggg  | gccagacggt  | gcagtcctgt  | gtcatcagga  | acctccggaa  | 4500 |
| gaaccccaac  | tttgacatct  | gcaccctctt  | catggaagtg  | atgctgccca  | gggaggagct  | 4560 |
| ctactgcccc  | cccatcaccg  | tcaaggtcat  | cgataaccgc  | cagtttggcc  | gccggcctgt  | 4620 |
| ggtgggcccag | tgtaccatcc  | gctccctgga  | gagcttccct  | tgtgaccctt  | actcggcgga  | 4680 |
| gagtcacatcc | ccacagggtg  | gcccagacga  | tgtgagccta  | ctcagtcctg  | gggaagacgt  | 4740 |
| gctcatcgac  | attgatgaca  | aggagcccct  | catccccatc  | caggaggaag  | agttcatcga  | 4800 |
| ttggtggagc  | aaattctttg  | cctccatagg  | ggagagggaa  | aagtgcggct  | cctacctgga  | 4860 |
| gaaggatttt  | gacaccctga  | aggtctatga  | cacacagctg  | gagaatgtgg  | aggcctttga  | 4920 |
| gggcctgtct  | gacttttgta  | acaccttcaa  | ctgtatccgg  | ggcaagacgc  | aggaggagac  | 4980 |
| agaagatcca  | tctgtgattg  | gtgaatttaa  | gggcctcttc  | aaaatttatc  | ccctcccaga  | 5040 |
| agaccagcc   | atccccatgc  | ccccaaagaca | gttccaccag  | ctggccgccc  | agggacccca  | 5100 |
| ggagtgcctg  | gtccgtatct  | acattgtccg  | agcatttggc  | ctgcagccca  | aggaccccaa  | 5160 |
| tggaaagtgt  | gactcttaca  | tcaagatctc  | catagggag   | aaatcagtga  | gtgaccagga  | 5220 |
| taactacatc  | ccctgcacgc  | tggagcccg   | atttggaag   | atgttcgagc  | tgacctgcac  | 5280 |
| tctgcctctg  | gagaaggacc  | taaagatcac  | tctctatgac  | tatgacctcc  | tctccaagga  | 5340 |
| cgaaaagatc  | ggtgagacgg  | tcgtcgacct  | ggagaacagg  | ctgctgtcca  | agtttggggc  | 5400 |
| tcgctgtgga  | ctcccacaga  | cctactgtgt  | ctctggaccg  | aaccagtggc  | gggaccagct  | 5460 |

15/68

|             |            |            |            |            |            |      |
|-------------|------------|------------|------------|------------|------------|------|
| ccgccccctcc | cagctcctcc | acctcttctg | ccagcagcat | agagtcaagg | cacctgtgta | 5520 |
| ccggacagac  | cgtgtaatgt | ttcaggataa | agaatattcc | attgaagaga | tagaggctgg | 5580 |
| caggatccca  | aaccacacac | tgggcccagt | ggaggagcgt | ctggctctgc | atgtgcttca | 5640 |
| gcagcagggc  | ctggtcccgg | agcacgtgga | gtcacggccc | ctctacagcc | ccctgcagcc | 5700 |
| agacatcgag  | caggggaagc | tgcagatgtg | ggtcgaccta | tttccgaagg | ccctggggcg | 5760 |
| gcctggacct  | cccttcaaca | tcacccacg  | gagagccaga | aggtttttcc | tgcgttgat  | 5820 |
| tatctggaat  | accagagatg | tgatcctgga | tgacctgagc | ctcacggggg | agaagatgag | 5880 |
| cgacatttat  | gtgaaagggt | ggatgattgg | ctttg      |            |            | 5915 |

<210> 4  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 4  
 tgggacctca aaggcatcc 20

<210> 5  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 5  
 accatgctgt aggatgtgga 20

<210> 6  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 6  
 gggaggtgaa gcaacttcaa 20

<210> 7  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 7  
 ctcacggggt agaagatgag 20

<210> 8  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 8  
 cagggccgag atgagccaa 20

<210> 9  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 9  
 acatcaaggg tcctggatga 20

<210> 10  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 10  
 ctgtggcggt gtttccggtg 20

<210> 11  
 <211> 20

16/68

<212> DNA  
<213> Homo sapiens

<400> 11  
acagacgtgc gttatcgttc

20

<210> 12  
<211> 20  
<212> DNA  
<213> Homo sapiens

<400> 12  
aagactgagc aaaatcccag

20

<210> 13  
<211> 6912  
<212> DNA  
<213> Homo sapiens

<400> 13  
tcgaccgccc agccaggtgc aaaatgccgt gtcattggga gactccgcag ccggagcatt 60  
agattacagc tcgacggagc tcgggaaggg cggcgggggg ggaagatgag cagaagcccc 120  
tgtttctcga acgcccggctg acaagcgggg tgagcgcagg cggggcgggg acccagccta 180  
gcccactgga gcagccgggg gtggcccgtt cccctttaag agcaactgct ctaagccagg 240  
agccagagat tcgagccggc ctgcgccagc cagccctctc cagcgagggg acccacaagc 300  
ggcgccctcg ccctcccagc ctttccgagc cctctttgcg ccctggggcg acggggccct 360  
acacgcgcca agcatgtctg gggctttcat cctctatgcc gagaacgtcc acacacccga 420  
caccgacatc agcgatgcct actgctccgc ggtgtttgca ggggtgaaga agagaaccaa 480  
agtcatacaag aacagcgtga accctgtatg gaatgagggg tttgaatggg acctcaaagg 540  
gcatccccct ggaccagggc tctgagcttc atgtggtggt caaagaccat gagacgatgg 600  
ggaggaacag gttcctgggg gaagccaagg tcccactccg agaggtcctc gccaccccta 660  
gtctgtccgc cagcttcaat gccccctgc tggacaccaa gaagcagccc acaggggcct 720  
cgctggctct gcaggtgtcc tacacaccgc tgccctggagc tgtgccccctg ttcccccccc 780  
ctactcctct ggagccctcc ccgactctgc ctggagatga tgtagtggca gacacaggag 840  
gagaggaaga cacagaggac cagggactca caaggaaact accttcacgt cctccgcccc 900  
aaagcggagg cccgggggct cccaccaccc caaggtacatc tagaaagctg ctgtcagaca 1020  
actaccocgg gatcaaaaaga aagcgaagtg cgcctacatc tgatcgaggg gcgccagctg ccgggggtga 1080  
aaccgcagga tttccagatc aggggtccagg cagggcgacg caagcggacg cggatccaca 1140  
acatcaagcc tgtggtcaag gttaccgctg ttttcttcaa cttgtttgac tctcctgggg 1200  
agggaaacag cccactcttc tttatcacgg ttgtagactc tcgttctctc aggacagatg 1260  
agctgtttga tgagcccatc atggacgtgg gcaccattta cagagagccc cggcacgcct 1320  
ctctcctcgg ggagttccgg ctctcagacc ctgatgactt ctctgctggg gccagaggct 1380  
atctcaggaa gtggctgctg gtgctggggc ctggggacga agcgccctctg gagagaaaag 1440  
acctgaaaac aagcctttgt gtgctggggc gcaacctgct ccggcccaca ggcgtagccc 1500  
acccctctga agacaaggag ctgaaggtct tccgggcccga ggacttgccg cagatggacg 1560  
tgcgaggagc ccacttctgc aaacagatct ttggcttcga gagtaacaag aagaacttgg 1620  
atgccgtgat tgtggaggtc agctttgctg ggaaaatgct gtgcagcaag atcttgaga 1680  
tggaacccct tgtggaggtc aaccagaaca tcacactgcc tgccatgttt ccctccatgt 1740  
agacggccaa ccctcagtgg atcatagact gggaccgcct gactcacaat gacatcgtgg 1800  
gcgaaaaaat gaggatctgt tcgaaaatct ctgcccctgg aggagaaata gaagaggagc 1860  
ctaccaccta cctgagtatg tgcgaaagcct cagacttgga tgactacctg ggcttcctcc 1920  
ctgcaggtgc tgtcaagcct atcaacctct atggcagtc cagagagttc acaggttcc 1980  
ccacttttgg gccctgctac aacacaggca aggggggaagg agcagagtga acagaaggtg cgtggccggc 2040  
cagaccctta cacagagctc aagctgggtg agcacagtga accttaggag gcgcaagtac tccctgtttg 2100  
ttctgtcttc cctggagacc gtggagaagt accttaggag atgtggatga tgccatccag tttgaggtca 2160  
ctgcggatga ctcagccacc atgtctgtagg tgacctgctt cccgctggcc tccaccactc 2220  
cgcccttcta ctacgggaac aagttcgaca actactacta aggacatcag ccatagaatc gagactcaga 2280  
gcatcggaag tgcagtcctt gacgggtgcc cctacctgga cctaccctgg ggtaacgtga 2340  
agtacagccg ggtgctgtca tccctatggg aggcagctga aagctggcct ggagcaggtc cacctggccc 2400  
aacctgtggt tgggattgct gaggacgtgg actcgctggt acatccatga gacaccctct gccaccacc 2460  
accagctgct tgggattgct gacgggtgcc cctctgggtg atcacctgag ccaaatactt gaggtgccc 2520  
tgaaggcgca gtgctccacg gaggacgtgg cctctgggtg cctgagctct ggagcagggc gaggtactggc 2580  
tcacgcagag ctgcagccag cctctgggtg cctgagctct cctgagctct cctgagctct atcgatctt 2640  
tggaacagta cctgtaccag agtgagctcc ctgcagctct cctgagctct cctgagctct atcgatctt 2700  
tgccctgaa gctcgccac agtgagctcc cctgagctct cctgagctct cctgagctct atcgatctt 2760  
tcctgctct gcgtgcccag gcagaggagc accagcgggt gcccggccac caagtctctt 2820  
ggatgctgca gggagacaag cgtgtggcat

|             |             |            |             |            |             |      |
|-------------|-------------|------------|-------------|------------|-------------|------|
| tctcccgccg  | gggtgccaac  | tactgtggca | agaattgtgg  | gaagctacag | acaatctttc  | 2880 |
| tgaatatatcc | gatggagaag  | gtgcctggcg | cccggatgcc  | agtgcagata | cggttcaagc  | 2940 |
| tgtggttttg  | gctctctgtg  | gatgagaagg | agttcaacca  | gtttgctgag | gggaagctgt  | 3000 |
| ctgtctttgc  | tgaacacctat | gagaacgaga | ctaagttggc  | ccttgttggg | aactggggca  | 3060 |
| caacgggcct  | cacctacccc  | aagttttctg | acgtcacggg  | caagatcaag | ctaccaagg   | 3120 |
| acagcttcgg  | cccctcgccc  | ggctggacct | gggctggaga  | ttggttcgtg | tgctccggaga | 3180 |
| agactctgct  | ccatgacatg  | gacgcccgtc | acctgagctt  | cgtggaagag | gtgtttgaga  | 3240 |
| accagaccgg  | gcttcccggg  | ggccagtggg | tctacatgag  | tgacaactac | accgatgtga  | 3300 |
| acggggagaa  | ggtgcttccc  | aaggatgaca | ttgagtggcc  | actgggctgg | aagtgggaag  | 3360 |
| atgaggaatg  | gtccacagac  | ctcaaccggg | ctgtcgatga  | gcaaggctgg | gagtatagca  | 3420 |
| tcaccatccc  | cccgagcg    | aagccgaagc | actgggtccc  | tgctgagaag | atgtactaca  | 3480 |
| cacaccgacg  | gcggcgctgg  | gtgcgcctgc | gcaggaggga  | tctcagccaa | atggaagcac  | 3540 |
| tgaaaaggca  | caggcaggcg  | gaggcggagg | gcgagggtcg  | ggagtacgcc | tctctttttg  | 3600 |
| gctggaagtt  | ccacctcgag  | taccgcaaga | cagatgcctt  | ccgcgcgcgc | cgctggcgcc  | 3660 |
| gtcgcgatgga | gccactggag  | aagacggggc | ctgcagctgt  | gtttgccctt | gagggggccc  | 3720 |
| tgggcgcgct  | gatggatgac  | aagagtgaag | attccatgtc  | cgtctccacc | ttgagcttcg  | 3780 |
| gtgtgaacag  | acccacgatt  | tcctgcatat | tcgactatgg  | gaaccgctac | catctacgct  | 3840 |
| gctacatgta  | ccaggccccg  | gacctggctg | cgatggacaa  | ggactctttt | tctgatccct  | 3900 |
| atgccatcgt  | ctccttccctg | caccagagcc | agaagacggt  | ggtggtgaag | aacaccctta  | 3960 |
| acccacacct  | ggaccagacg  | ctcatcttct | acgagatcga  | gatctttggc | gagccggcca  | 4020 |
| cagttgctga  | gcaaccgccc  | agcatttgtg | tggagctgta  | cgaccatgac | acttatggtg  | 4080 |
| cagacgagtt  | tatgggtcgc  | tgcactgtgc | aaccgagtct  | ggaacggatg | ccacggctgg  | 4140 |
| cctgggtccc  | actgacgagg  | ggcagccagc | cgtcggggga  | gctgctggcc | tcttttgagc  | 4200 |
| tcattccagag | agagaagccg  | gccatccacc | atattcctgg  | ttttgaggtg | caggagacat  | 4260 |
| caaggatcct  | ggtatgagtct | gaggacacag | acctgcctta  | cccaccaccc | cagaggagg   | 4320 |
| ccaacatcta  | catggttcct  | cagaacatca | agccagcgct  | ccagcgtacc | gccatcgaga  | 4380 |
| tcctggcatg  | gggcctgctg  | aacatgaaga | gttaccagct  | ggccaacatc | tcctccccc   | 4440 |
| gcctcgtggt  | agagtgtggg  | ggccagacgg | tgacgtcctg  | tgtcatcagg | aacctccgga  | 4500 |
| agaaccccaa  | ctttgacatc  | tgacccctct | tcattggaagt | gatgctgccc | agggaggagc  | 4560 |
| tctactgccc  | ccccatcacc  | gtcaaggcca | tcgataaccg  | ccagtttggc | cgccggcgctg | 4620 |
| tggtgggcca  | gtgtaccatc  | cgctccctgg | agagcttccct | gtgtgacccc | tactcgcgcg  | 4680 |
| agagtccatc  | cccacagggt  | ggcccagacg | atgtgagcct  | actcagtcct | ggggaagacg  | 4740 |
| tgctcatcga  | cattgatgac  | aaggagcccc | tcattcccat  | ccaggaggaa | gagttcatcg  | 4800 |
| attggtggag  | caaattcttt  | gcctccatag | gggagaggga  | aaagtgcggc | tcctactggg  | 4860 |
| agaaggatgt  | tgacacctg   | aaggtctatg | acacacagct  | ggagaatgtg | gaggcctttg  | 4920 |
| agggcctgtc  | tgacttttgt  | aacaccttca | agctgtaccg  | gggcaagacg | caggaggaga  | 4980 |
| cagaagatcc  | atctgtgatt  | ggtgaattta | agggcctctt  | caaaatttat | cccctcccag  | 5040 |
| aagaccacgc  | catccccatg  | cccccaagac | agttccacca  | gctggccgcc | cagggacccc  | 5100 |
| aggagtgcct  | ggtccgtatc  | tacattgtcc | gagcattttg  | cctgcagccc | aaggacccca  | 5160 |
| atggaaagtg  | tcctccttac  | atcaagatct | ccataggaaa  | gaaatcagtg | agtgaccagg  | 5220 |
| ataactacat  | cccctgcacg  | ctggagcccc | tatttgga    | gatgttcgag | ctgacctgca  | 5280 |
| ctctgcctct  | ggagaaggac  | ctaaagatca | ctctctatga  | ctatgacctc | ctctccaagg  | 5340 |
| acgaaaagat  | cggtgagacg  | gtcgtcgacc | tggagaacag  | gctgctgtcc | aagtttgggg  | 5400 |
| ctcgtctgtg  | actcccacag  | acctaactgt | tctctggacc  | gaaccagtgg | cggggaccagc | 5460 |
| tcgcgccctc  | ccagctcctc  | cacctcttct | gccagcagca  | tagagtcaag | gcacctgtgt  | 5520 |
| accggacaga  | ccgtgtaatg  | tttcaggata | agaatatttc  | cattgaagag | atagaggctg  | 5580 |
| gcaggatccc  | aaacccacac  | ctgggcccag | tggaggagcg  | tctggctctg | catgtgcttc  | 5640 |
| agcagcaggg  | cctgggtccc  | gagcacgtgg | agtcacggcc  | cctctacagc | cccctgcagc  | 5700 |
| cagacatcga  | gcagggggaag | ctgcagatgt | gggtcgacct  | atttccgaag | gccctggggc  | 5760 |
| ggcctggacc  | tccttccaac  | atcacccac  | ggagagccag  | aagggttttc | ctgcgttgta  | 5820 |
| ttatctggaa  | taccagagat  | gtgatcctgg | atgacctgag  | cctcacgggg | gagaagatga  | 5880 |
| gcgacattta  | tgtgaaaggt  | tggatgattg | gctttgaaga  | acacaagcaa | aagacagacg  | 5940 |
| tgcattatcg  | ttccctggga  | ggtgaaggca | acttcaactg  | gaggttcatt | ttccccttcg  | 6000 |
| actacctgce  | agctgagcaa  | gtctgtacca | ttgccaaagaa | ggatgccttc | tggaggctgg  | 6060 |
| acaagactga  | gagcaaaatc  | ccagcacgag | tgggtgttcca | gatctgggac | aatgacaagt  | 6120 |
| tctcctttga  | tgattttctg  | ggctccctgc | agctcgatct  | caaccgcatg | ccaagccag   | 6180 |
| ccaagacagc  | caagaagtgc  | tccttggacc | agctggatga  | tgctttccac | ccagaatggt  | 6240 |
| ttgtgtccct  | ttttgagcag  | aaaacagtga | agggctgggt  | gccctgtgta | gcagaagagg  | 6300 |
| gtgagaagaa  | aatactggcg  | ggcaagctgg | aaatgacctt  | ggagattgta | gcagagagtg  | 6360 |
| agcatgagga  | gcggcctgct  | ggccaggggc | gggatgagcc  | caacatgaac | cctaagcttg  | 6420 |
| aggacccaag  | gcgccccgac  | acctccttcc | tgtggtttac  | ctccccatac | aagaccatga  | 6480 |
| agttcatcct  | gtggcgcgct  | ttccgggtgg | ccatcatcct  | cttcatcatc | ctcttcatcc  | 6540 |
| tgtgtgctgt  | cctggccatc  | ttcatctacg | ccttcccga   | ctatgctgcc | atgaagctgg  | 6600 |
| tgaagccctt  | cagctgagga  | ctctcctggc | ctgtagaagg  | ggccgtgggg | tcccctccag  | 6660 |
| catgggactg  | gcctgcctcc  | tcgcccagc  | tcggcgagct  | cctccagacc | tcctaggcct  | 6720 |
| gattgtcctg  | ccagggtggg  | cagacagaca | gatggaccgg  | cccacactcc | cagagttgct  | 6780 |
| aacatggagc  | tctgagatca  | ccccacttcc | atcatttctt  | tctccccc   | cccaacgctt  | 6840 |



18/68

ttttggatca gctcagacat atttcagtat aaaacagttg gaaccacaaa aaaaaaaaaa 6900  
aaaaaaaaaa aa 6912

<210> 14  
<211> 6911  
<212> DNA  
<213> Homo sapiens

<400> 14  
tcgaccgccc agccaggtgc aaaatgccgt gtcattggga gactccgcag ccggagcatt 60  
agattacagc tcgacggagc tcgggaaggg cggcggggt ggaagatgag cagaagcccc 120  
tggtctcgga acgcccgtg acaagcgggg tgagcgcagg cggggcgggg acccagccta 180  
gcccactgga gcagccgggg gtggcccgtt cccctttaag agcaactgct ctaagccagg 240  
agccagagat tcgagccggc ctcgcccagc cagccctctc cagcgagggg acccacaagc 300  
ggcgccctcg ccctcccagc ctttccgagc cctctttgag ccctgggccc acggggccct 360  
acacgcgcca agcatgctga gggctctcat cctctatgcc gagaacgtcc acacaccgca 420  
caccgacatc agcgatgcct actgctccgc ggtgtttgca ggggtgaaga agagaaccaa 480  
agtcatacaag aacagcgtga accctgtatg gaatgagggg tttgaatggg acctcaaggg 540  
catccccctg gaccagggct ctgagcttca tgtggtgggt aaagaccatg agacgatggg 600  
gaggaacagg ttcctggggg aagccaaggt cccactccga gaggtcctcg ccaccctag 660  
tctgtccgcc agcttcaatg cccccgtct ggacaccaag aagcagccca caggggccc 720  
gctggtcctg caggtgtcct acacaccgct gcctggagct gtgcccctgt tcccgcctcc 780  
tactcctctg gagccctccc cgactctgcc tgacctggat gtagtggcag acacaggagg 840  
agaggaagac acagaggacc agggactcac tggagatgag gcggagccat tcctggatca 900  
aagcggaggg ccggggggct ccaccacccc aaggaaacta ccttcacgtc ctccgcccc 960  
ctaccccggg atcaaaaagaa agcgaagtgc gcctacatct agaaagctgc tgtcagacaa 1020  
accgcaggat ttccagatca ggggtccagg gatcgagggg cgccagctgc cgggggtgaa 1080  
catcaagcct gtggtcaagg ttaccgctgc agggcagacc aagcggagcg ggatccacaa 1140  
gggaaacagc ccactcttca atgagactct tttcttcaac ttgtttgact ctccgtggga 1200  
gctgtttgat gagcccatct ttatcacgtt ggtagactct cgttctctca ggacagatgc 1260  
tctcctcggg gagttccgga tggacgtggg caccatttac agagagcccc ggcacgccta 1320  
tctcaggaag tggctgctgc tctcagaccc tgatgacttc tctgctgggg ccagaggcta 1380  
cctgaaaaca agcctttgtg tgctggggcc tggggacgaa gcgcctctgg agagaaaaga 1440  
cccctctgaa gacaaggagg acattgaaag caacctgctc cggcccacag cgtagccct 1500  
gcgaggagcc cacttctgcc tgaaggtctt aacagatctt tggcttcgag agtaacaaga agaacttgg 1560  
tgccctgatg gacaacgtga aacagatctt gctttgcggg gaaaatgctg tgcagcaaga tcttgagaa 1620  
ggaccccttt gtggaggtca gctttgcggg cactctgctg gcccagctgc gccatgtttc cctccatgtg 1680  
gacggccaac cctcagtgga accagaacat cacactgcct ggaccgctg actcacaatg acatcggtg 1740  
cgaaaaaatg aggattcgta tcatagactg cgaaaatctc tgcccctgga ggagaaatag aagaggagcc 1800  
taccacctac ctgagtatgt gcaagcctc cgaaagcctc agacttggat gactacctgg gcttccctcc 1860  
tgacgtgctt ccctgtctca acacaggcaa agctggtgga gacactgccc agagagttca caggtctccc 1920  
agacccctac ctggagacca agctggtgga gcacagtga ccttaggagg tggtgatgat gccatccagt ccgctggcct 2040  
tgccgatgac atcctccggg tggagaagtg ccttaggagg tggtgatgat gccatccagt ccgctggcct 2040  
ggccttctac tcagccacca tgctgtagga gacctgcctg ctactactac ctacctgag catagaatcg agactcagaa 2100  
catcggaac tacgggaaca agttcgacat acgggtgcca cctactggga agctggcctg gagcaggtcc acctggccct 2160  
gtacagccgt gtgctgtcat cctactggga agctggcctg gctcagctga cggatgagct 2220  
acctgtggtg gggattgctg accggctgga ctcgctgggt catccatgag acaccctctg ccaccacct 2280  
gaaggcgcag tgctccacgg tgcagccagc ctgtaccagc ctggccaca gtgagctccc cagaggagcc 2340  
catcgaggc ctggagacca agctggtgga gcacagtga ccttaggagg tggtgatgat gccatccagt ccgctggcct 2400  
ggaccagtac ctggagacca agctggtgga gcacagtga ccttaggagg tggtgatgat gccatccagt ccgctggcct 2460  
ggccctgaag cctgctctg cgtgccctg ggagacaagc gtgtggcata ccagcgggtg gaattgtggg aagctacaga caatctttct 2520  
cctgctctg gatgctgcag ggtgccaaact actgtggcaa cgggatgcca gttcaaccag tttgctgagg cttgttggga actggggcac 2580  
gaaatatacc atggagaagg tgccctggcg cgggatgcca gttcaaccag tttgctgagg cttgttggga actggggcac 2640  
gtggtttggg ctctctgtgg gaaacctatg agaacgagac agttttctga cgtcacgggc ggtgagatg tgggtcgtgt 2700  
tgtctttgct aacctatg acctaccca ccctcggccg gctggacctg acgcccgtca gccagtggat ctacatgagt tgagtggcca ctgggctgga agtgaggaa 2760  
aacgggcctc acctaccca ccctcggccg gctggacctg acgcccgtca gccagtggat ctacatgagt tgagtggcca ctgggctgga agtgaggaa 2820  
cagcttccgc ccctcggccg gctggacctg acgcccgtca gccagtggat ctacatgagt tgagtggcca ctgggctgga agtgaggaa 2880  
gactctgctc catgacatgg cttcccggag aggatgacat tcaaccgggc ctgggctgga agtgaggaa 2940  
ccagaccggg gtgcttccca tccacagacc caggagcgga agccgaagca ctgggtccct gctgagaaga tgtactacac 3000  
cgggggagaag tgaggaatgg caccatcccc cggagcgga agccgaagca ctgggtccct gctgagaaga tgtactacac 3060  
tgaggaatgg caccatcccc cggagcgga agccgaagca ctgggtccct gctgagaaga tgtactacac 3120  
cgggggagaag tgaggaatgg caccatcccc cggagcgga agccgaagca ctgggtccct gctgagaaga tgtactacac 3180  
tgaggaatgg caccatcccc cggagcgga agccgaagca ctgggtccct gctgagaaga tgtactacac 3240  
cgggggagaag tgaggaatgg caccatcccc cggagcgga agccgaagca ctgggtccct gctgagaaga tgtactacac 3300  
tgaggaatgg caccatcccc cggagcgga agccgaagca ctgggtccct gctgagaaga tgtactacac 3360  
cgggggagaag tgaggaatgg caccatcccc cggagcgga agccgaagca ctgggtccct gctgagaaga tgtactacac 3420  
tgaggaatgg caccatcccc cggagcgga agccgaagca ctgggtccct gctgagaaga tgtactacac 3480

|             |             |             |            |            |            |      |
|-------------|-------------|-------------|------------|------------|------------|------|
| acaccgacgg  | cggcgctggg  | tgcgcctgcg  | caggagggat | ctcagccaaa | tggaagcact | 3540 |
| gaaaaggcac  | aggcaggcgg  | aggcggaggg  | cgagggctgg | gagtacgcct | ctcttttttg | 3600 |
| ctggaagtcc  | cacctcgagt  | accgcaagac  | agatgccttc | cgccgcgcgc | gctggcgccg | 3660 |
| tcgcatggag  | ccactggaga  | agacggggcc  | tgcagctgtg | tttgcccttg | agggggccct | 3720 |
| gggcggcgtg  | atggatgaca  | agagtgaaga  | ttccatgtcc | gtctccacct | tgagcttcgg | 3780 |
| tgtgaacaga  | cccacgattt  | cctgcatatt  | cgactatggg | aaccgctacc | atctacgtcg | 3840 |
| ctacatgtac  | caggccccgg  | acctggctgc  | gatggacaag | gactcttttt | ctgatcccta | 3900 |
| tgccatcgtc  | tccttcctgc  | accagagcca  | gaagacgggt | gtggtgaaga | acacccttaa | 3960 |
| ccccacctgg  | gaccagacgc  | tcatcttcta  | cgagatcgag | atctttggcg | agccggccac | 4020 |
| agttgctgag  | caaccgcccc  | gcattgtggt  | ggagctgtac | gaccatgaca | cttatgggtg | 4080 |
| agacgagttt  | atgggtcgct  | gcattctgtc  | accgagtctg | gaacggatgc | cacggctggc | 4140 |
| ctggttccca  | ctgacgaggg  | gcagccagcc  | gtcgggggag | ctgctggcct | cttttgagct | 4200 |
| catccagaga  | gagaagccgg  | ccatccacca  | tattcctggt | tttgaggtgc | aggagacatc | 4260 |
| aaggatcctg  | gatgagtctg  | aggacacaga  | cctgccctac | ccaccacccc | agagggaggc | 4320 |
| caacatctac  | atggttcctc  | agaacatcaa  | gccagcgctc | cagcgtaccg | ccatcgagat | 4380 |
| cctggcatgg  | ggcctgcgga  | acatgaagag  | ttaccagctg | gccaacatct | cctccccag  | 4440 |
| cctcgtggta  | gagtggtggg  | gccagacggg  | gcagtcctgt | gtcatcagga | acctccggaa | 4500 |
| gaaccccaac  | tttgacatct  | gcaccctcct  | catggaagtg | atgctgcca  | gggaggagct | 4560 |
| ctactgcccc  | cccatcaccc  | tcaaggtcat  | cgataaccgc | cagtttggcc | gccggcctgt | 4620 |
| ggtggggccg  | tgtaccatcc  | gctccctgga  | gagcttcctg | tgtgaccctt | actcgccgga | 4680 |
| gagtccatcc  | ccacaggggt  | gccagagcga  | tgtgagccta | ctcagtcctg | gggaagacgt | 4740 |
| gctcatcgac  | attgatgaca  | aggagcccct  | catccccatc | caggaggaag | agttcatcga | 4800 |
| ttggtggagc  | aaattctttg  | cctccatagg  | ggagagggaa | aagtgcggct | cctacctgga | 4860 |
| gaaggatttt  | gacaccctga  | aggtctatga  | cacacagctg | gagaatgtgg | aggcctttga | 4920 |
| gggcctgtct  | gacttttgtt  | acaccttcaa  | gctgtaccgg | ggcaagacgc | aggcaggagc | 4980 |
| agaagatcca  | tctgtgattg  | gtgaatttaa  | gggcctcttc | aaaattttat | ccctcccaga | 5040 |
| agaccagccc  | atccccatgc  | ccccaaagaca | gttccaccag | ctggccgccc | agggacccca | 5100 |
| ggagtgcctg  | gtccgtatct  | acattgtccg  | agcatttggt | ctgcagccca | aggaccccaa | 5160 |
| tggaaagtgt  | gatccttaca  | tcaagatctc  | catagggaag | aaatcagtga | gtgaccagga | 5220 |
| taactacatc  | ccctgcacgc  | tggagcccgt  | atttggaag  | atgttcgagc | tgacctgcac | 5280 |
| tctgcctctg  | gagaaggacc  | taaagatcac  | tctctatgac | tatgacctcc | tctccaagga | 5340 |
| cgaaaagatc  | ggtgagacgg  | tcgtcgacct  | ggagaacagg | ctgctgtcca | agtttggggc | 5400 |
| tcgctgtgga  | ctccacacaga | cctactgtgt  | ctctggaccg | aaccagtggc | gggaccagct | 5460 |
| ccgcccctcc  | cagctcctcc  | acctcttctg  | ccaagcagat | agagtcaagg | cacctgtgt  | 5520 |
| ccggacagac  | cggtcaatgt  | ttcaggataa  | agaatattcc | attgaagaga | tagaggctgg | 5580 |
| caggatccca  | aacccacacc  | tgggcccagt  | ggaggagcgt | ctggctctgc | atgtgcttca | 5640 |
| gcagcagggc  | ctggtcccgg  | agcacgtgga  | gtcacggccc | ctctacagcc | ccctgcagcc | 5700 |
| agacatcgag  | caggggaagc  | tgcagatgtg  | ggtcgacct  | tttccgaagg | ccctggggcg | 5760 |
| gcctggacct  | cccttcaaca  | tcacccacag  | gagagccaga | aggtttttcc | tgcgttgtat | 5820 |
| tatctggaat  | accagagatg  | tgatcctgga  | tgacctgagc | ctcacggggg | agaagatgag | 5880 |
| cgacatttat  | gtgaaagggt  | ggatgattgg  | ctttgaagaa | cacaagcaaa | agacagacgt | 5940 |
| gcattatcgt  | tccctgggag  | gtgaaggcaa  | cttcaactgg | aggttcattt | tccccttcga | 6000 |
| ctacctgcca  | gctgagcaag  | tctgtaccat  | tgccaagaag | gatgccttct | ggaggctgga | 6060 |
| caagactgag  | agcaaaatcc  | cagcacgagt  | ggtgttccag | atctgggaca | atgacaagtt | 6120 |
| ctcctttgat  | gattttctgg  | gctccctgca  | gctcgatctc | aaccgcatgc | ccaagccagc | 6180 |
| caagacagcc  | aagaagtgtc  | ccttggacca  | gctggatgat | gctttccacc | cagaatgggt | 6240 |
| tgtgtccctt  | tttgagcaga  | aaacagtga   | gggctgggtg | ccctgtgtag | cagaagaggg | 6300 |
| tgagaagaaa  | atactggcgg  | gcaagctgga  | aatgaccttg | gagattgtag | cagagagtga | 6360 |
| gcatgaggag  | cggcctgctg  | gccagggccg  | ggatgagccc | aacatgaacc | ctaagcttga | 6420 |
| ggacccaagg  | cgccccgaca  | cctccttcc   | gtggtttacc | tccccataca | agaccatgaa | 6480 |
| gttcacatctg | tggcggcggt  | tccggtgggc  | catcatcctc | ttcatcatcc | tcttcacatc | 6540 |
| gctgctgttc  | ctggccatct  | tcatctacgc  | cttcccgaac | tatgctgcca | tgaagctggg | 6600 |
| gaagcccttc  | agctgaggac  | tctcctgccc  | tgtagaaggg | gccgtggggg | cccctccagc | 6660 |
| atgggactgg  | cctgctcct   | ccgcccagct  | cggcgagctc | ctccagacct | cctaggcctg | 6720 |
| attgtcctgc  | cagggtgggc  | agacagacag  | atggaccggc | ccacactccc | agagttgcta | 6780 |
| acatggagct  | ctgagatcac  | cccacttcca  | tcatctcctt | ctcccccaac | ccaacgcttt | 6840 |
| tttgatcag   | ctcagacata  | tttcagtata  | aaacagttgg | aaccacaaaa | aaaaaaaaaa | 6900 |
| aaaaaaaaaa  | a           |             |            |            |            | 6911 |

&lt;210&gt; 15

&lt;211&gt; 6910

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 15

|            |            |            |            |            |            |     |
|------------|------------|------------|------------|------------|------------|-----|
| tcgaccgccc | agccaggtgc | aaaatgccgt | gtcattggga | gactccgcag | ccggagcatt | 60  |
| agattacagc | tcgacggagc | tcgggaaggg | cggcgggggg | ggaagatgag | cagaagcccc | 120 |

20/68

|             |             |            |             |             |            |      |
|-------------|-------------|------------|-------------|-------------|------------|------|
| tgttctcggg  | acgcccggctg | acaagcgggg | tgagcgcagg  | cggggcgggg  | accagccta  | 180  |
| gcccactgga  | gcagccgggg  | gtggcccgtt | cccccttaag  | agcaactgct  | ctaagccagg | 240  |
| agccagagat  | tcgagccggc  | ctcgcccagc | cagccctctc  | cagcgagggg  | accacaagc  | 300  |
| ggcgccctcg  | ccctcccgac  | ctttccgagc | cctcttttgcg | ccctgggccc  | acggggccct | 360  |
| acacgcgcca  | agcatgctga  | gggtcttcat | cctctatgcc  | gagaacgtcc  | acacaccoga | 420  |
| caccgacatc  | agcgatgcct  | actgctccgc | gggtgttgca  | ggggtgaaga  | agagaaccaa | 480  |
| agtcatcaag  | aacagcgtga  | accctgtatg | gaatgagggg  | tttgaatggg  | acctcaagg  | 540  |
| catccccctg  | gaccagggct  | ctgagcttca | tgtgggtggc  | aaagaccatg  | agacgatggg | 600  |
| gaggaacagg  | ttcctggggg  | aagccaaggt | cccactccga  | gaggtcctcg  | ccacccttag | 660  |
| tctgtccggc  | agcttcaatg  | ccccctgct  | ggacaccaag  | aagcagccca  | caggggcctc | 720  |
| gctggtcctg  | caggtgtcct  | acacaccgct | gcctggagct  | gtgcccctgt  | tcccgcctcc | 780  |
| tactcctctg  | gagccctccc  | cgactctgcc | tgacctggat  | gtagtggcag  | acacaggagg | 840  |
| agaggaagac  | acagaggacc  | agggactcac | tggagatgag  | gcggagccat  | tcctggatca | 900  |
| aagcggaggc  | ccggggggctc | ccaccacccc | aaggaaacta  | ccttcacgtc  | ctccgcccc  | 960  |
| ctaccccggg  | atcaaaaaga  | agcgaagtgc | gcctacatct  | agaaaagtgc  | tgtcagacaa | 1020 |
| accgcaggat  | ttccagatca  | gggtccaggt | tgatcgaggg  | cgccagctgc  | cgggggtgaa | 1080 |
| catcaagcct  | gtggtcaagg  | ttaccgctgc | agggcagacc  | aagcggacgc  | ggatccacaa | 1140 |
| gggaaacagc  | ccactcttca  | atgagactct | tttcttcaac  | ttgtttgact  | ctcctgggga | 1200 |
| gctgtttgat  | gagcccatct  | ttatcacggt | ggtagactct  | cgcttctctca | ggacagatgc | 1260 |
| tctcctcggg  | gagttccgga  | tggacgtggg | caccatttac  | agagagcccc  | ggcacgccta | 1320 |
| tctcaggaag  | tggtctctgc  | tctcagaccc | tgatgacttc  | tctgctgggg  | ccagaggcta | 1380 |
| cctgaaaaca  | agcctttgtg  | tgctggggcc | tggggacgaa  | gcgcctctgg  | agagaaaaga | 1440 |
| ccccctctgaa | gacaaggagg  | acattgaaag | caacctgctc  | cggcccacag  | gcgtagccct | 1500 |
| gcgaggagcc  | cacttctgcc  | tgaaggtctt | ccgggcccag  | gacttgccgc  | agatggacga | 1560 |
| tgccgtgatg  | gacaacgtga  | aacagatctt | tggcttcgag  | agtaacaaga  | agaacttggt | 1620 |
| ggaccccttt  | gtggaggtca  | gctttgcggg | gaaaatgctg  | tgacgcaaga  | tcttggagaa | 1680 |
| gacggccaac  | ctctagtgga  | accagaacat | cacactgcct  | gccatgtttc  | cctccatgtg | 1740 |
| cgaaaaaatg  | aggattcgtg  | tcatagactg | ggaccgcctg  | actcacaatg  | acatcgtggc | 1800 |
| taccacctac  | ctgagtatgt  | cgaaaatctc | tgcccctgga  | ggagaaatag  | aagaggagcc | 1860 |
| tgcaggtgct  | gtcaagcctt  | cgaaagcctc | agacttgatg  | gactacctgg  | gcttctctcc | 1920 |
| cactttttggg | ccctgtctaca | tcaacctcta | tggcagtcct  | agagagttca  | cagggtctcc | 1980 |
| agacccttac  | acagagctca  | acacaggcaa | gggggaaggt  | gtggcttatc  | gtggccggct | 2040 |
| tctgctctcc  | ctggagacca  | agctgggtga | gcacagtga   | cagaaggtgg  | aggaccttcc | 2100 |
| tgccggatgac | atcctccggg  | tggagaagta | ccttaggagg  | cgcaagtact  | ccctgtttgc | 2160 |
| ggcctttctac | tcagccacca  | tgctgcagga | tgtggatgat  | gccatccagt  | ttgaggtcag | 2220 |
| catcgggaac  | tacgggaaca  | agttcgacat | gcctgcctcg  | ccgctggcct  | ccaccactca | 2280 |
| gtacagccgt  | gcagtctttg  | acgggtgcca | ctactactac  | ctaccctggg  | gtaacgtgaa | 2340 |
| acctgtgggtg | gtgctgtcat  | cctactggga | ggacatcagc  | catagaatcg  | agactcagaa | 2400 |
| ccagctgctt  | gggattgtctg | accggctgga | agctggcctg  | gagcaggtcc  | acctggccct | 2460 |
| gaaggcgag   | tgctccacgg  | aggacgtgga | ctcgctgggt  | gctcagctga  | cggatgagct | 2520 |
| catcgacagg  | ctcgaccagc  | ctctgggtga | catccatgag  | acaccctctg  | ccaccacct  | 2580 |
| ggaccagtac  | ctgtaccagc  | tgccgaccca | tcacctgagc  | caaactactg  | aggctgcct  | 2640 |
| ggccctgaag  | ctcgccacca  | gtgagctccc | tgacgctctg  | gagcaggcgg  | aggactggct | 2700 |
| cctgcgtctg  | cgtgccctgg  | cagaggagcc | ccagaacagc  | ctgccggaca  | tcgtcatctg | 2760 |
| gatgctgac   | ggagacaagc  | gtgtggcata | ccagcgggtg  | cccgccacc   | aagtcctctt | 2820 |
| ctcccggcgg  | ggtgccaact  | actgtggcaa | gaattgtggg  | aagctacaga  | caatctttct | 2880 |
| gaaatatccg  | atggagaagg  | tgcttgccgc | ccgatgcca   | gtgcagatac  | gggtcaagct | 2940 |
| gtgggtttggg | ctctctgtgg  | atgagaagga | gttcaaccag  | tttgcctgag  | ggaagctgtc | 3000 |
| tgtctttgct  | gaaacctatg  | agaacgagac | taagttggcc  | cttgttggga  | actggggcac | 3060 |
| aacgggcctc  | acctacccca  | agttttctga | cgctcacggg  | aagatcaagc  | tacccaagga | 3120 |
| cagcttccgc  | ccctcgcccg  | gctggacctg | ggctggagat  | tggttcgtgt  | gtccggagaa | 3180 |
| gactctgctc  | catgacatgg  | acgccggtca | cctgagcttc  | gtggaagagg  | tgtttgagaa | 3240 |
| ccagaccogg  | cttcccggag  | gccagtggat | ctacatgagt  | gacaactaca  | ccgatgtgaa | 3300 |
| cggggagaag  | gtgcttccca  | aggatgacat | tgagtgccta  | ctgggctgga  | agtgggaaga | 3360 |
| tgaggaatgg  | ttccacagacc | tcaaccgggc | tgtcgatgag  | caaggctggg  | agtatagcat | 3420 |
| caccatcccc  | ccggagcggg  | agccgaagca | ctgggtccct  | gctgagaaga  | tgtactacac | 3480 |
| acaccgacgg  | cggcgctggg  | tgccgctgcg | caggagggat  | ctcagccaaa  | tggaagcact | 3540 |
| gaaaaggcac  | aggcaggcgg  | aggcggaggg | cgagggctgg  | gagtacgcct  | ctcttttttg | 3600 |
| ctggaagtgc  | cacctcgagt  | accgcaagac | agatgccttc  | cgccgcgcgc  | gctggcccg  | 3660 |
| tcgcatggag  | ccactggaga  | agacggggcc | tgacagctgtg | tttgcccttg  | agggggccct | 3720 |
| gggcggcgtg  | atggatgaca  | agagtgaaga | ttccatgtcc  | gtctccacct  | tgagcttcgg | 3780 |
| tgtgaacaga  | cccacgattt  | cctgcatatt | cgactatggg  | aaccgctacc  | atctacgctg | 3840 |
| ctacatgtac  | caggcccggg  | acctggctgc | gatggacaag  | gactcttttt  | ctgacctcta | 3900 |
| tgccatcgtc  | tccttcctgc  | accagagcca | gaagacgggt  | gtggtgaaga  | acacctttaa | 3960 |
| ccccacctgg  | gaccagacgc  | tcatcttcta | cgagatcgag  | atctttggcg  | agccggccac | 4020 |
| agttgctgag  | caaccgcccc  | gcatttgtgt | ggagctgtac  | gaccatgaca  | cttatggtgc | 4080 |
| agacgagttt  | atgggtcgct  | gcatttgtca | accgagctctg | gaacggatgc  | cacggctggc | 4140 |

21/68

|            |             |             |             |             |             |      |
|------------|-------------|-------------|-------------|-------------|-------------|------|
| ctggttccca | ctgacgaggg  | gcagccagcc  | gtcgggggag  | ctgctggcct  | cttttgagct  | 4200 |
| catccagaga | gagaagccgg  | ccatccacca  | tattcctggt  | tttgaggtgc  | aggagacatc  | 4260 |
| aaggatcctg | gatgagtctg  | aggacacaga  | cctgccctac  | ccaccacccc  | agagggaggc  | 4320 |
| caacatctac | atggttcctc  | agaacatcaa  | gccagcgtc   | cagcgtaccg  | ccatcgagat  | 4380 |
| cctggcatgg | ggcctgcgga  | acatgaagag  | ttaccagctg  | gccaacatct  | cctccccag   | 4440 |
| cctcgtggtg | gagtgtgggg  | gccagacggt  | gcagtcctgt  | gtcatcagga  | acctccggaa  | 4500 |
| gaaccccaac | tttgacatct  | gcacctctt   | catggaagt   | atgctgcca   | gggaggagct  | 4560 |
| ctactgcccc | cccatcaccg  | tcaaggtcat  | cgataaccgc  | cagtttgccc  | gccggcctgt  | 4620 |
| gggtgggccg | tgtaccatcc  | gctccctgga  | gagcttcctg  | tgtgaccctt  | actcggcgga  | 4680 |
| gagtccatcc | ccacagggtg  | gcccagacga  | tgtgagccca  | ctcagtcctg  | gggaagacgt  | 4740 |
| gctcatcgac | attgatgaca  | aggagcccc   | catccccatc  | caggaggaag  | agttcatcga  | 4800 |
| ttggtggagc | aaattctttg  | cctccatagg  | ggagagggaa  | aagtgcggct  | cctacctgga  | 4860 |
| gaaggatttt | gacaccctga  | aggtctatga  | cacacagctg  | gagaatgtgg  | aggcctttga  | 4920 |
| gggcctgtct | gacttttgta  | acaccttcaa  | gctgtaccgg  | ggcaagacgc  | aggaggagac  | 4980 |
| agaagatcca | tctgtgattg  | gtgaatttaa  | gggcctcttc  | aaaattttatc | ccctcccaga  | 5040 |
| agaccctgac | atccccatgc  | cccccaagaca | gttccaccag  | ctggccgccc  | agggacccca  | 5100 |
| ggagtgcctg | gtccgtatct  | acattgtccg  | agcatttgge  | ctgcagccca  | aggaccccaa  | 5160 |
| tggaaagtgt | gaccccttaca | tcaagatctc  | catagggaag  | aaatcagtg   | gtgaccagga  | 5220 |
| taactacatc | ccctgcacgc  | tggagcccg   | atttggaag   | atgttcgagc  | tgacctgcac  | 5280 |
| tctgcctctg | gagaaggacc  | taaagatcac  | tctctatgac  | tatgacctcc  | tctccaagga  | 5340 |
| cgaaaagatc | ggtgagacgg  | tcgtcgacct  | ggagaacagg  | ctgctgtcca  | agtttggggc  | 5400 |
| tcgctgtgga | ctccacacaga | cctactgtgt  | ctctggaccg  | aaccagtggc  | gggaccagct  | 5460 |
| ccgccccctc | cagctcctcc  | acctcttctg  | ccagcagcat  | agagtcaagg  | cacctgtgta  | 5520 |
| ccggacagac | cgtgtaatgt  | ttcaggataa  | agaatattcc  | attgaagaga  | tagaggctgg  | 5580 |
| caggatccca | aacccacacc  | tgggcccag   | ggaggagcgt  | ctggctctgc  | atgtgcttca  | 5640 |
| gcagcagggc | ctggctccgg  | agcacgtgga  | gtcacggccc  | ctctacagcc  | ccctgcagcc  | 5700 |
| agacatcgag | caggggaagc  | tgcagatgtg  | ggctcgacct  | tttccgaagg  | ccctggggcg  | 5760 |
| gcctggacct | cccttcaaca  | tcaccccacg  | gagagccaga  | aggtttttcc  | tgcgttgtat  | 5820 |
| tatctggaat | accagagatg  | tgatcctgga  | tgacctgagc  | ctcacggggg  | agaagatgag  | 5880 |
| cgacatttat | gtgaaagggt  | ggatgattgg  | ctttgaagaa  | cacaagcaaa  | agacagacgt  | 5940 |
| gcattatcgt | tccctgggag  | gtgaagcaac  | ttcaactgga  | ggttcatttt  | ccccttcgac  | 6000 |
| tacctgccag | ctgagcaagt  | ctgtaccatt  | gccaagaagg  | atgccttctg  | gaggctggac  | 6060 |
| aagactgaga | gcaaaatccc  | agcacgagtg  | gtgttccaga  | tctgggacaa  | tgacaagttc  | 6120 |
| tcctttgatg | atthttctggg | ctccctgcag  | ctcgatctca  | accgcagtc   | caagccagcc  | 6180 |
| aagacagcca | agaagtgtc   | cttggaaccg  | ctggatgatg  | ctttccaccc  | agaatgggtt  | 6240 |
| gtgtccagtt | ttgagcagaa  | aacagtgaag  | ggctgggtgg  | cctgtgtagc  | agaagagggg  | 6300 |
| gagaagaaaa | tactggcggg  | caagctggaa  | atgaccttgg  | agattgtagc  | agagagttag  | 6360 |
| catgaggagc | ggcctgctgg  | ccaggggccg  | gatgagccca  | acatgaaccc  | taagcttgag  | 6420 |
| gacccaaggc | gccccgacac  | ctccttcctg  | tggtttacct  | ccccatacaa  | gaccatgaag  | 6480 |
| ttcatcctgt | ggcgcggttt  | ccggtgggcc  | atcatcctct  | tcatacctct  | cttcatcctg  | 6540 |
| ctgctgttcc | tgcccatctt  | catctacgcc  | ttcccgaaact | atgctgccat  | gaagctgggtg | 6600 |
| aagcccttca | gctgaggact  | ctcctgccct  | gtagaagggg  | ccgtgggggtc | ccctccagca  | 6660 |
| tgggactggc | ctgcctcctc  | cgcccagctc  | ggcgagctcc  | tccagacctc  | ctaggcctga  | 6720 |
| ttgtcctgcc | aggggtgggg  | gacagacaga  | tggaccggcc  | cacactccca  | gagttgttaa  | 6780 |
| catggagctc | ccacttcacc  | ccattctctc  | aatctctctc  | tcccccaacc  | caacgctttt  | 6840 |
| ttggatcagc | tcagacatat  | ttcagtataa  | aacagttgga  | accacaaaaa  | aaaaaaaaaa  | 6900 |
| aaaaaaaaaa |             |             |             |             |             | 6910 |

&lt;210&gt; 16

&lt;211&gt; 6911

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 16

|             |            |            |            |            |            |     |
|-------------|------------|------------|------------|------------|------------|-----|
| tcgaccgccc  | agccaggtgc | aaaatgccgt | gtcattggga | gactccgcag | ccggagcatt | 60  |
| agattacagc  | tcgacggagc | tcgggaagg  | cgccgggggt | ggaagatgag | cagaagcccc | 120 |
| tgttctcgga  | acgccggctg | acaagcggg  | tgagcgcagg | cggggcgggg | accagcccta | 180 |
| gcccactgga  | gcagccgggg | gtggcccggt | cccccttaag | agcaactgct | ctaagccagg | 240 |
| agccagagat  | tcgagccggc | ctcgcccgac | cagccctctc | cagcgagggg | accacaagc  | 300 |
| ggcgccctcg  | ccctcccgac | ctttccgagc | cctctttgcg | ccctgggcgc | acggggccct | 360 |
| acacgcgcca  | agcatgctga | gggtcttcat | cctctatgcc | gagaacgtcc | acacaccgga | 420 |
| caccgacatc  | agcgatgcct | actgctccgc | gggtgttgca | ggggtgaaga | agagaaccaa | 480 |
| agtcataaag  | aacagcgtga | accctgtatg | gaatgagggg | tttgaatggg | acctcaaggg | 540 |
| catccccctg  | gaccagggct | ctgagcttca | tgtggtggtc | aaagaccatg | agacgatggg | 600 |
| gaggaacagg  | ttcttggggg | aagccaaggt | ccactccga  | gaggtcctcg | ccacccttag | 660 |
| tctgtccgcc  | agcttcaatg | ccccctgct  | ggacaccaag | aagcagccca | caggggcctc | 720 |
| gctgggtcctg | caggtgtcct | acacaccgct | gcctggagct | gtgccccctg | tcccgccccc | 780 |

|             |             |             |             |            |             |      |
|-------------|-------------|-------------|-------------|------------|-------------|------|
| tactcctctg  | gagccctccc  | cgactctgcc  | tgacctggat  | gtagtggcag | acacaggagg  | 840  |
| agaggaagac  | acagaggacc  | agggactcac  | tggagatgag  | gcggagccat | tcctggatca  | 900  |
| aagcggaggc  | ccgggggctc  | ccaccacccc  | aaggaaacta  | ccttcacgtc | ctccgcccc   | 960  |
| ctaccccggg  | atcaaaaagaa | agcgaagtgc  | gcctacatct  | agaaagctgc | tgctagacaa  | 1020 |
| accgcaggat  | ttccagatca  | gggtccagg   | gatcgagggg  | cgccagctgc | cgggggtgaa  | 1080 |
| catcaagcct  | gtggtcaagg  | ttaccgctgc  | agggcagacc  | aagcggacgc | ggatccacaa  | 1140 |
| gggaaacagc  | ccactcttca  | atgagactct  | tttcttcaac  | ttgtttgact | ctcctgggga  | 1200 |
| gctgtttgat  | gagcccatct  | ttatcacggt  | ggtagactct  | cgttctctca | ggacagatgc  | 1260 |
| tctcctcggg  | gagttccgga  | tggacgtggg  | caccattttac | agagagcccc | ggcacgccta  | 1320 |
| tctcaggaag  | tggctgctgc  | tctcagaccc  | tgatgacttc  | tctgctgggg | ccagaggcta  | 1380 |
| cctgaaaaca  | agcctttgtg  | tgctggggcc  | tggggacgaa  | gcgcctctgg | agagaaaaga  | 1440 |
| cccctctgaa  | gacaaggagg  | acattgaaag  | caacctgctc  | cggcccacag | gcgtagccct  | 1500 |
| gcgaggagcc  | cactttctgcc | tgaaggtctt  | ccgggcccag  | gacttgccgc | agatggacga  | 1560 |
| tgccgtgatg  | gacaacgtga  | aacagatctt  | tggcttcgag  | agtaacaaga | agaacttggt  | 1620 |
| ggaccccttt  | gtggaggtca  | gctttgcggg  | gaaaatgctg  | tgcagcaaga | tcttggaaga  | 1680 |
| gacggccaac  | cctcagtgga  | accagaacat  | cacactgcct  | gccatgtttc | cctccatgtg  | 1740 |
| cgaaaaaatg  | aggattcgta  | tcatagactg  | ggaccgcctg  | actcacaatg | acatcgtggc  | 1800 |
| taccacctac  | ctgagtatgt  | cgaaaatctc  | tgcccttgga  | ggagaaatag | aagaggagcc  | 1860 |
| tgcaggtgct  | gtcaagcctt  | cgaaagcctc  | agacttggtg  | gactacctgg | gcttctctcc  | 1920 |
| cacttttggg  | ccctgctaca  | tcaacctcta  | tggcagctcc  | agagagttca | caggcttccc  | 1980 |
| agacccctac  | acagagctca  | acacaggcaa  | gggggaaggt  | gtggcttata | gtggccggct  | 2040 |
| tctgctctcc  | ctggagacca  | agctggtgga  | gcacagtga   | cagaaggtgg | aggaccttcc  | 2100 |
| tgccgatgac  | atcctccggg  | tggagaagta  | ccttaggagg  | cgcaagtact | ccctgtttgc  | 2160 |
| ggccttctac  | tcagccacca  | tgctgcagga  | tgtggatgat  | gccatccagt | ttgaggtcag  | 2220 |
| catcggaac   | tacgggaaca  | agttcgacat  | gacctgcctg  | ccgctggcct | ccaccactca  | 2280 |
| gtacagccgt  | gcagtctttg  | acgggtgcca  | ctactactac  | ctaccctggg | gtaacgtgaa  | 2340 |
| acctgtgggt  | gtgctgtcat  | cctactggga  | ggacatcagc  | catagaatcg | agactcagaa  | 2400 |
| ccagctgctt  | gggattgctg  | accggtgga   | agctggcctg  | gagcaggtcc | acctggccct  | 2460 |
| gaaggcgag   | tgctccacgg  | aggacgtgga  | ctcgtggtg   | gctcagctga | cggatgagct  | 2520 |
| catcgaggg   | tgcagccagc  | ctctgggtga  | catccatgag  | acaccctctg | ccaccacact  | 2580 |
| ggaccagtac  | ctgtaccagc  | tgccgaccga  | tcacctgagc  | caaatactg  | aggctgccct  | 2640 |
| ggcctgaag   | ctcgccca    | gtgagctccc  | tgacgtctg   | gagcaggcgg | aggactggct  | 2700 |
| cctgctctg   | cgtgccctgg  | cagaggagcc  | ccagaacagc  | ctgccggaca | tcgtcatctg  | 2760 |
| gatgctgcag  | ggagacaagc  | gtgtggcata  | ccagcgggtg  | cccgccacc  | aagtcctctt  | 2820 |
| ctcccggcgg  | ggtgccaa    | actgtggcaa  | gaattgtggg  | aagctacaga | caatctttct  | 2880 |
| gaaatattccg | atggagaagg  | tgccgtggcg  | ccggatgcc   | gtgcagatac | gggtcaagct  | 2940 |
| gtggtttggg  | ctctctgtgg  | atgagaagga  | gttcaaccag  | tttgctgagg | ggaagctgtc  | 3000 |
| tgtctttgct  | gaaacctatg  | agaacgagac  | taagtggcc   | cttggtggga | actggggcac  | 3060 |
| aacgggcctc  | acctacccca  | agttttctga  | cgtcacgggc  | aagatcaagc | taccacaagg  | 3120 |
| cagcttccgc  | ccctcggccg  | gctggacctg  | ggctggagat  | tggttcgtgt | gtccggagaa  | 3180 |
| gactctgctc  | catgacatgc  | acgcccgtca  | cctgagcttc  | gtggaagagg | tgtttgagaa  | 3240 |
| ccagaccggg  | cttcccggag  | gccagtggat  | ctacatgagt  | gacaactaca | ccgatgtgaa  | 3300 |
| cggggagaag  | gtgcttccca  | aggatgacat  | tgagtgcaca  | ctgggctgga | agtgggaaga  | 3360 |
| tgaggaatgg  | tccacagacc  | tcaaccgggc  | tgctgatgag  | caaggctggg | agtatagcat  | 3420 |
| caccatcccc  | ccggagcgga  | agccgaagca  | ctgggtccct  | gctgagaaga | tgtactacac  | 3480 |
| acaccgacgg  | cggcgctggg  | tgccgctgcg  | caggagggat  | ctcagccaaa | tggagcact   | 3540 |
| gaaaaggcac  | aggcaggcgg  | aggcggaggg  | cgagggtgg   | gagtacgcct | ctcttttttg  | 3600 |
| ctggaagtgc  | cacctcgagt  | accgcaagac  | agatgccttc  | cgcgcgcggc | gctggcgccg  | 3660 |
| tcgcatggag  | ccactggaga  | agacggggcc  | tgacgtgtg   | tttgcccttg | agggggccct  | 3720 |
| gggcgcgctg  | atggatgaca  | agagtgaaga  | ttccatgtcc  | gtctccacct | tgagcttcgg  | 3780 |
| tgtgaacaga  | cccacgattt  | cctgcatatt  | cgactatggg  | aaccgctacc | atctacgctg  | 3840 |
| ctacatgtac  | caggcccggg  | acctggctgc  | gatggacaag  | gactcttttt | ctgatcccta  | 3900 |
| tgccatcgtc  | tccttcctgc  | accagagcca  | gaagacgggt  | gtggtgaaga | acacccttaa  | 3960 |
| ccccacctgg  | gaccagacgc  | tcactctcta  | cgagatcgag  | atctttggcg | agccggccac  | 4020 |
| agttgtctgag | caaccgcccc  | gcattgtggt  | ggagctgtac  | gaccatgaca | cttatggtgc  | 4080 |
| agacgagttt  | atgggtcgct  | gcactctgtca | accgagtctg  | gaacggatgc | cacggctggc  | 4140 |
| ctgggtccca  | ctgacgaggg  | gcagccagcc  | gtcgggggag  | ctgctggcct | cttttgagct  | 4200 |
| catccagaga  | gagaagccgg  | ccatccacca  | tattcctggt  | tttgaggtgc | aggagacatc  | 4260 |
| aaggatcctg  | gatgagtctg  | aggacaacaga | cctgccttac  | ccaccacccc | agagggaggc  | 4320 |
| caacatctac  | atggttcctc  | agaacatcaa  | gccagcgctc  | cagcgtaccg | ccatcgagat  | 4380 |
| cctggcatgg  | ggcctgcgga  | acatgaagag  | ttaccagctg  | gccaacatct | cctccccag   | 4440 |
| cctcgtggta  | gagtgtgggg  | gccagacggt  | gcagtccctg  | gtcatcagga | acctccggaa  | 4500 |
| gaaccccaac  | tttgacatct  | gcaccctctt  | catggaagtg  | atgctgcca  | gggagagct   | 4560 |
| ctactgcccc  | cccataccg   | tcaaggtcat  | cgataaccgc  | cagtttgcc  | gcccggcctgt | 4620 |
| ggtgggcccag | gttaccatcc  | gctccctgga  | gagcttccctg | tgtgacccct | actcggcgga  | 4680 |
| gagtccatcc  | ccacagggtg  | gcccagacga  | tgtgagccta  | ctcagtcctg | gggaagacgt  | 4740 |
| gctcatcgac  | attgatgaca  | aggagccctt  | catccccatc  | caggaggaag | agttcatcga  | 4800 |

23/68

|            |             |             |             |             |            |      |
|------------|-------------|-------------|-------------|-------------|------------|------|
| ttggtggagc | aaattctttg  | cctccatagg  | ggagagggaa  | aagtgcggct  | cctacctgga | 4860 |
| gaaggatttt | gacaccctga  | aggtctatga  | cacacagctg  | gagaatgtgg  | aggcctttga | 4920 |
| gggcctgtct | gactttttgta | acaccttcaa  | gctgtaccgg  | ggcaagacgc  | aggaggagac | 4980 |
| agaagatcca | tctgtgattg  | gtgaatttaa  | gggcctcttc  | aaaatttatc  | ccctcccaga | 5040 |
| agaccagcc  | atccccatgc  | ccccaaagaca | gttccaccag  | ctggccgccc  | agggacccca | 5100 |
| ggagtgtctg | gtccgtatct  | acattgtccg  | agcatttggc  | ctgcagccca  | aggaccccaa | 5160 |
| tggaaagtgt | gacccctaca  | tcaagatctc  | catagggaaag | aaatcagtga  | gtgaccagga | 5220 |
| taactacatc | ccctgcacgc  | tggagcccgt  | atgttgaaag  | atgttcgagc  | tgacctgcac | 5280 |
| tctgcctctg | gagaaggacc  | taaagatcac  | tctctatgac  | tatgacctcc  | tctccaagga | 5340 |
| cgaaaagatc | ggtgagacgg  | tcgtcgacct  | ggagaacagg  | ctgctgtcca  | agtgtggggc | 5400 |
| tcgctgtgga | ctcccacaga  | cctactgtgt  | ctctggaccg  | aaccagtggc  | gggaccagct | 5460 |
| cgcgccctcc | cagctcctcc  | acctcttctg  | ccagcagcat  | agagtcaagg  | cacctgtgta | 5520 |
| ccggacagac | cgtgtaatgt  | ttcaggataa  | agaatattcc  | attgaagaga  | tagaggctgg | 5580 |
| caggatccca | aaccacacac  | tgggcccagt  | ggaggagcgt  | ctggctctgc  | atgtgcttca | 5640 |
| gcagcagggc | ctggtcccgg  | agcacgtgga  | gtcacggccc  | ctctacagcc  | ccctgcagcc | 5700 |
| agacatcgag | caggggaagc  | tgcatgtgtg  | gttcgacctc  | tttccgaagg  | ccctggggcg | 5760 |
| gcctggacct | cccttcaaca  | tcaccccacg  | gagagccaga  | agggttttcc  | tgctgtgtat | 5820 |
| tatctggaat | accagagatg  | tgatcctgga  | tgacctgagc  | ctcacggggg  | agaagatgag | 5880 |
| cgacatttat | gtgaaaggtt  | ggatgattgg  | ctttgaagaa  | cacaagcaaa  | agacagacgt | 5940 |
| gcattatcgt | tccctgggag  | gtgaaggcaa  | cttcaactgg  | aggttcattt  | tccccttcga | 6000 |
| ctacctgcca | gctgagcaag  | tctgtaccat  | tgccaagaag  | gatgccttct  | ggaggctgga | 6060 |
| caagactgag | agcaaaatcc  | cagcacgagt  | ggtgttccag  | atctggggaca | atgacaagtt | 6120 |
| ctcctttgat | gattttctgg  | gctccctgca  | gctcgatctc  | aaccgcatgc  | ccaagccagc | 6180 |
| caagacagcc | aagaagtgtc  | ccttggacca  | gctggatgat  | gctttccacc  | cagaatgggt | 6240 |
| tgtgtccctt | tttgagcaga  | aaacagtga   | gggctgggtg  | ccctgtgtag  | cagaagaggg | 6300 |
| tgagaagaaa | atactggcgg  | gcaagctgga  | aatgaccttg  | gagattgtag  | cagagagtga | 6360 |
| gcatgaggag | cggcctgtctg | gccagggccg  | ggatgagccc  | aacatgaacc  | ctaagcttga | 6420 |
| ggacccaagg | cgccccgaca  | cctccttcc   | gtgggtttacc | tccccataca  | agaccatgaa | 6480 |
| gttcatcctg | tggcggcggt  | tccgggtggg  | catcatcctc  | ttcatcatcc  | tcttcatcct | 6540 |
| gctgctgttc | ctggccatct  | tcactctacg  | cttcccgaac  | tatgctgcca  | tgaagctggg | 6600 |
| gaagcccttc | agctgaggac  | tctcctgccc  | tgtagaaggg  | gccgtggggg  | cccctccagc | 6660 |
| atgggactgg | cctgcctcct  | ccgcccagct  | cggcgagctc  | ctccagacct  | cctaggcctg | 6720 |
| attgtcctgc | caggggtggg  | agacagacag  | atgggaccgg  | ccacactccc  | agagttgcta | 6780 |
| acatggagct | ctgagatcac  | cccacttcca  | tcatttccct  | ctcccccaac  | ccaacgcttt | 6840 |
| tttggatcac | ctcagacata  | tttcagtata  | aaacagttgg  | aaccacaaaa  | aaaaaaaaaa | 6900 |
| aaaaaaaaaa | a           |             |             |             |            | 6911 |

&lt;210&gt; 17

&lt;211&gt; 6911

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 17

|             |            |             |             |            |             |      |
|-------------|------------|-------------|-------------|------------|-------------|------|
| tcgaccgccc  | agccaggtgc | aaaatgccgt  | gtcattggga  | gactccgcag | ccggagcatt  | 60   |
| agattacagc  | tcgacggagc | tcgggaaggg  | cgccgggggt  | ggaagatgag | cagaagcccc  | 120  |
| tgttctcgga  | acgccggctg | acaagcgggg  | tgagcgcagg  | cggggccccg | accagcccta  | 180  |
| gcccactgga  | gcagccgggg | gtggcccgtt  | cccctttaag  | agcaactgct | ctaagccagg  | 240  |
| agccagagat  | tcgagccggc | ctcgcccagc  | cagccctctc  | cagcgagggg | accacaagc   | 300  |
| ggcgctctcg  | ccctcccagc | ctttccgagc  | cctctttgcg  | ccctggggcg | acggggccct  | 360  |
| acacgcgcca  | agcatgctga | gggtcttcat  | cctctatgcc  | gagaacgtcc | acacaccgga  | 420  |
| caccgacatc  | agcgatgcct | actgctccgc  | ggtgtttgca  | ggggtgaaga | agagaaccaa  | 480  |
| agtcacaaag  | aacagcgtga | accctgtatg  | gaatgaggga  | tttgaatggg | acctcaaggg  | 540  |
| catccccctg  | gaccagggct | ctgagcttca  | tgtggtggtc  | aaagaccatg | agacgatggg  | 600  |
| gaggaacagg  | ttcctggggg | aagccaaggt  | ccactccga   | gaggtcctcg | ccacccttag  | 660  |
| tctgtccgcc  | agcttcaatg | ccccctgct   | ggacaccaag  | aagcagccca | caggggcctc  | 720  |
| gctggtcctg  | caggtgtcct | acacaccgct  | gcctggagct  | gtgcccctgt | tcccggcccc  | 780  |
| tactcctctg  | gagccctccc | cgactctgcc  | tgacctggat  | gtagtggcag | acacaggagg  | 840  |
| agaggaagac  | acagaggacc | agggactcac  | tggagatgag  | gcggagccat | tcctggatca  | 900  |
| aagcggaggc  | ccgggggctc | ccaccacccc  | aaggaaacta  | ccttcacgtc | ctccgccccca | 960  |
| ctaccccggg  | atcaaaagaa | agcgaaagtgc | gcctacatct  | agaaagctgc | tgtcagacaa  | 1020 |
| accgcaggat  | ttccagatca | gggtccaggt  | gatcgagggg  | cgccagctgc | cggggggtgaa | 1080 |
| catcaagcct  | gtggtcaagg | ttaccgctgc  | agggcagacc  | aagcggacgc | ggatccacaa  | 1140 |
| gggaaacagc  | ccactcttca | atgagactct  | tttcttcaac  | ttgtttgact | ctcctgggga  | 1200 |
| gctggtttgat | gagcccatct | ttatcacggt  | ggtagactct  | cggtctctca | ggcacagatgc | 1260 |
| tctcctcggg  | gagttccgga | tggacgtggg  | caccattttac | agagagcccc | ggcacgccta  | 1320 |
| tctcaggaag  | tggctgctgc | tctcagaccc  | tgatgacttc  | tctgctgggg | ccagaggcta  | 1380 |
| cctgaaaaca  | agcctttgtg | tgctggggcc  | tggggacgaa  | gcgcctctgg | agagaaaaga  | 1440 |

24/68

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| ccccctctgaa | gacaaggagg  | acattgaaag  | caacctgctc  | cggcccacag  | gcgtagccct  | 1500 |
| gcgaggagcc  | cacttctgcc  | tgaaggtctt  | ccgggcccag  | gacttgccgc  | agatggacga  | 1560 |
| tgccgtgatg  | gacaacgtga  | aacagatctt  | tggcttcgag  | agtaacaaga  | agaacttggt  | 1620 |
| ggaccccttt  | gtggaggtca  | gctttgcggg  | gaaaatgctg  | tgcagcaaga  | tcttgagaa   | 1680 |
| gacggccaac  | cctcagtgga  | accagaacat  | cacactgcct  | gccatgtttc  | cctccatgtg  | 1740 |
| cgaaaaaatg  | aggattcgta  | tcatagactg  | ggaccgcctg  | actcacaatg  | acatcgtggc  | 1800 |
| taccacctac  | ctgagtatgt  | cgaaaatctc  | tgccccctgga | ggagaaatag  | aagaggagcc  | 1860 |
| tgcaggtgct  | gtcaagcctt  | cgaaagcctc  | agacttggtg  | gactacctgg  | gcttcctccc  | 1920 |
| cacttttggg  | ccctgctaca  | tcaacctcta  | tggcagtgccc | agagagttca  | cagggtctccc | 1980 |
| agacccctac  | acagagctca  | acacaggcaa  | gggggaaggt  | gtggcttatc  | gtggccggct  | 2040 |
| tctgctctcc  | ctggagacca  | agctggtgga  | gcacagtga   | cagaaggtgg  | aggaccttcc  | 2100 |
| tgcggtgac   | atcctccggg  | tggagaagta  | ccttaggagg  | cgcaagtact  | ccctgtttgc  | 2160 |
| ggccttctac  | tcagccacca  | tgtgcagga   | tgtggatgat  | gccatccagt  | ttgaggtcag  | 2220 |
| catcggaac   | tacgggaaca  | agttcgacat  | gacctgcctg  | ccgctggcct  | ccaccactca  | 2280 |
| gtacagccgt  | gcagtccttg  | acgggtgcca  | ctactactac  | ctaccctggg  | gtaacgtgaa  | 2340 |
| acctgtgggt  | gtgctgtcat  | cctactggga  | ggacatcagc  | catagaatcg  | agactcagaa  | 2400 |
| ccagctgctt  | gggattgctg  | accggctgga  | agctggcctg  | gagcaggtcc  | acctggccct  | 2460 |
| gaaggcgag   | tgtctccacg  | aggacgtgga  | ctcgctgggt  | gctcagctga  | cggatgagct  | 2520 |
| catcgaggg   | tgcagccagc  | ctctgggtga  | catccatgag  | acaccctctg  | ccaccacct   | 2580 |
| ggaccagtac  | ctgtaccagc  | tgcgcaccca  | tcacctgagc  | caaactcactg | aggctgccct  | 2640 |
| ggccctgaag  | ctcgccacca  | gtgagctccc  | tgcagctctg  | gagcaggcgg  | aggactggct  | 2700 |
| cctgcgtctg  | cgtgccctgg  | cagaggagcc  | ccagaacagc  | ctgccggaca  | tcgtcatctg  | 2760 |
| gatgctgcag  | ggagacaagc  | gtgtggcata  | ccagcgggtg  | cccgccacc   | aagtcctctt  | 2820 |
| ctcccgcg    | ggtgccaaact | actgtggcaa  | cggtatgcca  | aagctacaga  | caatctttct  | 2880 |
| gaaatatccg  | atggagaagg  | tgcttgccgc  | gttcaaccag  | tttgctgagg  | gggtcaagct  | 2940 |
| gtggtttggg  | ctctctgtgg  | atgagaagga  | taagttggcc  | cttggtggga  | ggaagctgtc  | 3000 |
| tgtctttgct  | gaaacctatg  | agaacgagac  | cgtcacgggc  | aagatcaagc  | actggggcac  | 3060 |
| aacgggcctc  | acctacccca  | agttttctga  | ggctggagat  | tggttcgtgt  | tacccaagga  | 3120 |
| cagcttccgc  | ccctcgcccg  | gctggacctg  | cctgagcttc  | gtgggaagagg | gtccggagaa  | 3180 |
| gactctgctc  | catgacatgg  | acgcccgtgc  | ctacatgagt  | gtggttctgt  | tggtttgagaa | 3240 |
| ccagaccggg  | cttcccggag  | gccagtggat  | tgagtgccca  | gacaactaca  | ccgatgtgaa  | 3300 |
| cggggagaga  | gtgcttccca  | aggatgacat  | tgtcgatgag  | ctgggctgga  | agtgggaaga  | 3360 |
| tgaggaatgg  | tccacagacc  | tcaaccgggc  | ctgggtccct  | caaggctggg  | agtatagcat  | 3420 |
| caccatcccc  | ccggagcgga  | agccgaagca  | caggagggat  | gctgagaaga  | tgactacac   | 3480 |
| acaccgacgg  | cggcgctggg  | tgcgcctgcg  | cgagggtgg   | ctcagccaaa  | tggaagcact  | 3540 |
| gaaaaggcac  | aggcaggcgg  | aggcgagggg  | agatgccttc  | gagtagcctc  | ctctttttgg  | 3600 |
| ctggaagtgc  | cacctcgagt  | accgcaagac  | tgcagctgtg  | cgccgcggcc  | gctggcgccg  | 3660 |
| tcgcatggag  | ccactggaga  | agacggggcc  | ttccatgtcc  | tttgcccttg  | agggggccct  | 3720 |
| gggcggcggt  | atggatgaca  | agagtgaaga  | ggactgtctc  | gtctccacct  | tgagcttcgg  | 3780 |
| tgtgaacaga  | cccacgattt  | cctgcatatt  | gactatggg   | aaccgctacc  | atctacgctg  | 3840 |
| ctacatgtac  | caggcccggg  | acctggctgc  | gatggacaag  | gactcttttt  | ctgatcccta  | 3900 |
| tgccatcgtc  | tccttctctg  | accagagcca  | gaagacgggt  | gtggtgaaga  | acacccttaa  | 3960 |
| ccccacctgg  | gaccagacgc  | tcattcttcta | cgagatcgag  | atctttggcg  | agccggccac  | 4020 |
| agttgctgag  | caaccgcccc  | gcattgtggt  | ggagctgtac  | gacctgaca   | cttatgggtg  | 4080 |
| agacgagttt  | atgggtcgct  | gcattgttca  | accgagctct  | gaacggatgc  | cacggctggc  | 4140 |
| ctggttccca  | ctgacgaggg  | gcagccagcc  | gtcgggggag  | ctgctggcct  | cttttgagct  | 4200 |
| catccagaga  | gagaagccgg  | ccatccacca  | tattcctggt  | tttgaggtgc  | aggagacatc  | 4260 |
| aaggatcctg  | gatgagctct  | aggacacaga  | cctgccctac  | ccaccacccc  | agaggagggc  | 4320 |
| caacatctac  | atggttcctc  | agaacatcaa  | gccagcgctc  | cagcgtaccg  | ccatcgagat  | 4380 |
| cctggcatgg  | ggcctgcgga  | acatgaagag  | ttaccagctg  | gccaacatct  | cctccccag   | 4440 |
| cctcgtggta  | gagtggtggg  | gccagacggg  | gcagtcctgt  | gtcatcagga  | acctccggaa  | 4500 |
| gaaccccaac  | tttgacatct  | gcacctctt   | catggaagtg  | atgctgcccc  | gggaggagct  | 4560 |
| ctactgcccc  | cccatacccg  | tcaaggtcat  | cgataaccgc  | cagtttgccc  | gccggcctgt  | 4620 |
| ggtgggcccag | tgtaccatcc  | gctccctgga  | gagcttcctg  | tgtgacctct  | actcggcgga  | 4680 |
| gagtcacatcc | ccacagggtg  | gccagacga   | tgtgagccta  | ctcagtcctg  | gggaagacgt  | 4740 |
| gctcatcgac  | attgatgaca  | aggagcccct  | catccccatc  | caggaggaag  | agttcatcga  | 4800 |
| ttggtggagc  | aaattctttg  | cctccatagg  | ggagaggga   | aagtgcggct  | cctaccttga  | 4860 |
| gaaggatttt  | gacaccctga  | aggtctatga  | cacacagctg  | gagaatgtgg  | aggcctttga  | 4920 |
| gggcctgtct  | gacttttgta  | acaccttcaa  | gctgtaccgg  | ggcaagacgc  | aggaggagac  | 4980 |
| agaagatcca  | tctgtgattg  | gtgaatttaa  | gggcctcttc  | aaaattttatc | ccctcccaga  | 5040 |
| agaccagcc   | atccccatgc  | ccccaaagaca | gttccaccag  | ctggccgccc  | agggacccca  | 5100 |
| ggagtgcctg  | gtccgtatct  | acattgtccg  | agcatttggc  | ctgcagccca  | aggaccccaa  | 5160 |
| tggaaagtgt  | gatccttaca  | tcaagatctc  | cataggaag   | aaatcagtga  | gtgaccagga  | 5220 |
| taactacatc  | ccctgcacgc  | tggagcccg   | atttggaaa   | atgttcgagc  | tgacctgcac  | 5280 |
| tctgcctctg  | gagaaggacc  | taaagatcac  | tctctatgac  | tatgacctcc  | tctccaagga  | 5340 |
| cgaaaagatc  | ggtgagacgg  | tcgtcgacct  | ggagaacagg  | ctgctgtcca  | agtttggggc  | 5400 |
| tcgctgtgga  | ctcccacaga  | cctactgtgt  | ctctggaccg  | aaccagtggc  | gggaccagct  | 5460 |



25/68

|             |             |             |            |            |            |      |
|-------------|-------------|-------------|------------|------------|------------|------|
| ccgccccctcc | cagctcctcc  | acctcttctg  | ccagcagcat | agagtcaagg | cacctgtgta | 5520 |
| ccggacagac  | cgtgtaatgt  | ttcaggataa  | agaatattcc | attgaagaga | tagaggctgg | 5580 |
| caggatccca  | aaccacacac  | tgggcccagt  | ggaggagcgt | ctggctctgc | atgtgcttca | 5640 |
| gcagcagggc  | ctgggtcccgg | agcacgtgga  | gtcacggccc | ctctacagcc | ccctgcagcc | 5700 |
| agacatcgag  | caggggaagc  | tgcagatgtg  | ggtcgacctt | tttccgaagg | ccctggggcg | 5760 |
| gcctggacct  | cccttcaaca  | tcacccacg   | gagagccaga | aggtttttcc | tgcgttgtat | 5820 |
| tatctggaat  | accagagatg  | tgatcctgga  | tgacctgagc | ctcacggggg | agaagatgag | 5880 |
| cgacatttat  | gtgaaagggt  | ggatgattgg  | ctttgaagaa | cacaagcaaa | agacagacgt | 5940 |
| gcattatcgt  | tccctgggag  | gtgaaggcaa  | cttcaactgg | aggttcattt | tccccctcga | 6000 |
| ctacctgcca  | gctgagcaag  | tctgtacctt  | tgccaagaag | gatgccttct | ggaggctgga | 6060 |
| caagactgag  | agcaaaatcc  | cagcacgagt  | gggtgtccag | atctgggaca | atgacaagtt | 6120 |
| ctcctttgat  | gattttctgg  | gctccctgca  | gctcgatctc | aaccgcatgc | ccaagccagc | 6180 |
| caagacagcc  | aagaagtgtc  | ccttggacca  | gctggatgat | gctttccacc | cagaatgggt | 6240 |
| tgtgtccctt  | tttgagcaga  | aaacagtga   | gggctgggtg | ccctgtgtag | cagaagaggg | 6300 |
| tgagaagaaa  | atactggcgg  | gcaagctgga  | aatgaccttg | gagattgtag | cagagactga | 6360 |
| gcatgaggag  | cggcctgctg  | gccagggccg  | agatgagccc | aacatgaacc | ctaagcttga | 6420 |
| ggacccaagg  | cgccccgaca  | cctccttcc   | gtggtttacc | tccccataca | agaccatgaa | 6480 |
| gttcatcctg  | tggcggcggt  | tccggtgggc  | catcatcctc | ttcatcatcc | tcttcatcct | 6540 |
| gctgctgttc  | ctggccatct  | tcatctacgc  | cttcccgaac | tatgctgcca | tgaagctgg  | 6600 |
| gaagcccttc  | agctgaggac  | tctcctgccc  | tgtagaaggg | gccgtggggg | cccctccagc | 6660 |
| atgggacttg  | cctgcctcct  | ccgcccagct  | cggcgagctc | ctccagacct | cctaggcctg | 6720 |
| attgtcctgc  | caggggtggc  | agacagacag  | atggaccggc | ccacactccc | agagttgcta | 6780 |
| acatggagct  | ctgagatcac  | cccacttcca  | tcatttcctt | ctcccccaac | ccaacgcttt | 6840 |
| tttggatcag  | ctcagacata  | tttcagttata | aaacagttgg | aaccacaaaa | aaaaaaaaaa | 6900 |
| aaaaaaaaaa  | a           |             |            |            |            | 6911 |

&lt;210&gt; 18

&lt;211&gt; 6911

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 18

|             |             |             |            |            |             |      |
|-------------|-------------|-------------|------------|------------|-------------|------|
| tcgaccgccc  | agccaggtgc  | aaaatgccgt  | gtcattggga | gactccgcag | ccggagcatt  | 60   |
| agattacagc  | tcgacggagc  | tcgggaagg   | cggcgggggg | ggaagatgag | cagaagcccc  | 120  |
| tgttctcgga  | acgccggctg  | acaagcgggg  | tgagcgcagg | cggggcgggg | ccccagccta  | 180  |
| gcccactgga  | gcagccgggg  | gtggcccgtt  | cccccttaag | agcaactgct | ctaagccagg  | 240  |
| agccagagat  | tcgagccggc  | ctcgcccagc  | cagccctctc | cagcgagggg | accacacaagc | 300  |
| ggcgcctcgg  | ccctcccagc  | ctttccgagc  | cctctttgcg | ccctggggcg | acggggccct  | 360  |
| acacgcgcca  | agcatgtctga | gggtcttcat  | cctctatgcc | gagaacgtcc | acacacccca  | 420  |
| caccgacatc  | agcgtgcctt  | actgctccgc  | gggtgttgca | ggggtgaaga | agagaaccaaa | 480  |
| agtcatcaag  | aacagcgtga  | accctgtatg  | gaatgaggga | tttgaatggg | acctcaaggg  | 540  |
| catccccctg  | gaccagggct  | ctgagcttca  | tgtgggtggc | aaagaccatg | agacgatggg  | 600  |
| gaggaacagg  | ttcctggggg  | aagccaaggt  | cccactccga | gaggtcctcg | ccacccctag  | 660  |
| tctgtccgcc  | agcttcaatg  | ccccctgct   | ggacaccaag | aagcagccca | caggggcctc  | 720  |
| gctgggtcctg | caggtgtcct  | acacaccgct  | gctggagct  | gtgcccctgt | tcccccccc   | 780  |
| tactcctctg  | gagccctccc  | cgactctgcc  | tgacctggat | gtagtggcag | acacaggagg  | 840  |
| agaggaagac  | acagaggacc  | agggactcac  | tggagatgag | gcggagccat | tcctggatca  | 900  |
| aagcggaggc  | ccgggggctc  | ccaccacccc  | aaggaaacta | ccttcacgtc | ctccgccccca | 960  |
| ctaccccggg  | atcaaaaaga  | agcgaagtgc  | gcctacatct | agaaagctgc | tgtcagacaa  | 1020 |
| accgcaggat  | ttccagatca  | gggtccaggt  | gatcgagggg | cgccagctgc | cgggggtgaa  | 1080 |
| catcaagcct  | gtgggtcaagg | ttaccgctgc  | agggcagacc | aagcggacgc | ggatccacaa  | 1140 |
| gggaaacagc  | ccactcttca  | atgagactct  | tttcttcaac | ttgtttgact | ctcctgggga  | 1200 |
| gctgtttgat  | gagcccattc  | ttatcacggg  | ggtagactct | cgttctctca | ggacagatgc  | 1260 |
| tctcctcggg  | gagttccgga  | tggacgtggg  | caccatttac | agagagcccc | ggcacgccta  | 1320 |
| tctcaggaag  | tggctgctgc  | tctcagaccc  | tgatgacttc | tctgctgggg | ccagaggcta  | 1380 |
| cctgaaaaca  | agcctttgtg  | tgtgggggcc  | tggggacgaa | gcgcctctgg | agagaaaaga  | 1440 |
| cccctctgaa  | gacaaggagg  | acattgaaag  | caacctgctc | cggcccacag | gcgtagccct  | 1500 |
| gcgaggagcc  | cacttctgcc  | tgaagggtct  | ccgggcccag | gacttgccgc | agatggacga  | 1560 |
| tgccgtgatg  | gacaacgtga  | aacagatcct  | ggtcttcgag | agtaacaaga | agaacttggt  | 1620 |
| ggaccccttt  | tgggaggtca  | gctttgcggg  | gaaaatgctg | tgacgcaaga | tcttgagaaa  | 1680 |
| gacggccaac  | cctcagtgga  | accagaacat  | cacactgcct | gccatgtttc | cctccatgtg  | 1740 |
| cgaaaaaatg  | aggattcgta  | tcatagactg  | ggaccgcctg | actcacaatg | acatcgtggc  | 1800 |
| taccacctac  | ctgagtatgt  | cgaaaatctc  | tggccctgga | ggagaaaatg | aagagagacc  | 1860 |
| tgcaggtgct  | gtcaagcctt  | cgaaagcctc  | agacttggat | gactacctgg | gcttcctccc  | 1920 |
| cacttttggg  | ccctgctaca  | tcaacctcta  | tggcagtcct | agagagttca | caggcttccc  | 1980 |
| agacccttac  | acagagctca  | acacaggcaa  | gggggaagg  | gtggcttatc | gtggccggct  | 2040 |
| tctgctctcc  | ctggagacca  | agctgggtgga | gcacagtga  | cagaaggtgg | aggaccttcc  | 2100 |



|             |             |             |            |             |             |      |
|-------------|-------------|-------------|------------|-------------|-------------|------|
| tgcggatgac  | atcctccggg  | tggagaagta  | ccttaggagg | cgcaagtact  | ccctgtttgc  | 2160 |
| ggccttttac  | tcagccacca  | tgctgcagga  | tgtggatgat | gccatccagt  | ttgaggtcag  | 2220 |
| catcgggaac  | tacgggaaca  | agttcgacat  | gacctgcctg | ccgctggcct  | ccaccactca  | 2280 |
| gtacagccgt  | gcagtctttg  | acgggtgcca  | ctactactac | ctaccctggg  | gtaacgtgaa  | 2340 |
| acctgtggtg  | gtgctgtcat  | cctactggga  | ggacatcagc | catagaatcg  | agactcagaa  | 2400 |
| ccagctgctt  | gggattgctg  | accggctgga  | agctggcctg | gagcaggtcc  | acctggccct  | 2460 |
| gaaggcgagc  | tgctccacgg  | aggacgtgga  | ctcgtggtg  | gctcagctga  | cggatgagct  | 2520 |
| catcgccagg  | tgccagccagc | ctctgggtga  | catccatgag | acaccctctg  | ccaccacact  | 2580 |
| ggaccagtac  | ctgtaccagc  | tgcgcaccca  | tcacctgagc | caaatcactg  | aggctgcctt  | 2640 |
| ggccctgaag  | ctcggccaca  | gtgagctccc  | tgcagctctg | gagcaggcgg  | aggactggct  | 2700 |
| cctgctgctg  | cgtgccctgg  | cagaggagcc  | ccagaacagc | ctgcccggaca | tcgtcatctg  | 2760 |
| gatctgagc   | ggagacaagc  | gtgtggcata  | ccagcgggtg | cccgcccacc  | aagtccctctt | 2820 |
| ctcccgcg    | gggtccaact  | actgtggcaa  | gaattgtggg | aagctacaga  | caatctttct  | 2880 |
| gaaatatccg  | atggagaagg  | tgcctggcgc  | ccggatgcca | gtgcagatac  | gggtcaagct  | 2940 |
| gtggtttggg  | ctctctgtgg  | atgagaagga  | gttcaaccag | tttgctgagg  | ggaagctgtc  | 3000 |
| tgtctttgct  | gaaacctatg  | agaacgagac  | taagttggcc | cttggtggga  | actggggcac  | 3060 |
| aacgggcctc  | acctaaccaca | agttttctga  | cgtcacgggc | aagatcaagc  | tacccaagga  | 3120 |
| cagcttccgc  | ccctcggccg  | gctggacctg  | ggctggagat | tggttcgtgt  | gtccggagaa  | 3180 |
| gactctgctc  | catgacatgg  | acgccggtca  | cctgagcttc | gtggaagagg  | tgtttgagaa  | 3240 |
| ccagaccccg  | cttcccggag  | gccagtggtg  | ctacatgagt | gacaactaca  | ccgatgtgaa  | 3300 |
| cggggagaa   | gtgcttccca  | aggatgacat  | tgagtggcca | ctgggctgga  | agtgggaaga  | 3360 |
| tgaggaatgg  | tccacagacc  | tcaaccgggc  | tgctgatgag | caaggctggg  | agtatagcat  | 3420 |
| caccatcccc  | ccggagcgga  | agccgaagca  | ctgggtccct | gctgagaaga  | tgtactacac  | 3480 |
| acaccgacgg  | cggcgctggg  | tgcgcctgcg  | caggagggat | ctcagccaaa  | tggaagcact  | 3540 |
| gaaaaggcac  | aggcaggcgg  | aggcggaggg  | cgagggtctg | gagtacgcct  | ctctttttgg  | 3600 |
| ctggaagtgc  | cacctcgagt  | accgcaagac  | agatgccttc | cgccgccgcc  | gctggcgccg  | 3660 |
| tcgcatggag  | ccactggaga  | agacggggcc  | tgcagctgtg | tttgcccttg  | agggggccct  | 3720 |
| gggcgcgctg  | atggatgaca  | agagtgaaga  | ttccatgtcc | gtctccacct  | tgagctctcg  | 3780 |
| tgtgaacaga  | cccacgattt  | cctgcatatt  | cgactatggg | aaccgctacc  | atctacgtg   | 3840 |
| ctacatgtac  | caggcccggg  | acctggctgc  | gatggacaag | gactcttttt  | ctgatcccta  | 3900 |
| tgccatcgte  | tccttccctg  | accagagcca  | gaagacgggt | gtggtgaaga  | acacccttaa  | 3960 |
| cccacactgg  | gaccagacgc  | tcattcttcta | cgagatcgag | atctttggcg  | agccggccac  | 4020 |
| agttgctgag  | caaccgcca   | gcattgtggt  | ggagctgtac | gacctgaca   | cttatgggtg  | 4080 |
| agacgagttt  | atgggtcgct  | gcatctgtca  | accgagctct | gaacggatgc  | cacgggtggc  | 4140 |
| ctggttccca  | ctgacgaggg  | gcagccagcc  | gtcgggggag | ctgctggcct  | cttttgagct  | 4200 |
| catccagaga  | gagaagccgg  | ccatccacca  | tattcctggt | tttgaggtgc  | aggagacatc  | 4260 |
| aagggtcctg  | gatgagtctg  | aggacacaga  | cctgccctac | ccaccacccc  | agagggaggc  | 4320 |
| caacatctac  | atgggtccct  | agaacatcaa  | gccagcgctc | cagcgtaccg  | ccatccagat  | 4380 |
| cctggcatgg  | ggcctgcgga  | acatgaagag  | ttaccagctg | gccaacatct  | cctccccag   | 4440 |
| cctcgtggta  | gagtggtggg  | gccagacggg  | gcagtcctgt | gtcatcagga  | acctccggaa  | 4500 |
| gaaccccaac  | tttgacatct  | gcacctctt   | catggaagtg | atgctgcca   | gggaggagct  | 4560 |
| ctactgcccc  | cccatacccg  | tcaaggctcat | cgataaccgc | cagtttgggc  | gccggcctgt  | 4620 |
| ggtgggcccag | tgtaccatcc  | gctccctgga  | gagcttccct | tgtgaccctt  | actcggcgga  | 4680 |
| gagtccatcc  | ccacagggtg  | gcccagacga  | tgtagcccta | ctcagtcctg  | gggaagacgt  | 4740 |
| gctcatcgac  | attgatgaca  | catccccatc  | ggagagggaa | caggaggaag  | agttcatcga  | 4800 |
| ttggtggagc  | aaattctttg  | cctccatagg  | cacacagctg | aagtgcggct  | cctacctgga  | 4860 |
| gaaggatttt  | gacaccctga  | aggtctatga  | gctgtaccgg | gagaatgtgg  | aggcctttga  | 4920 |
| gggcctgtct  | gacttttgta  | acaccttcaa  | gggcctcttc | ggcaagacgc  | aggaggagac  | 4980 |
| agaagatcca  | tctgtgattg  | gtgaatttaa  | gttccaccag | aaaattttatc | ccctcccaga  | 5040 |
| agacccagcc  | atccccatgc  | ccccaaagaca | agcatttggc | ctggccgccc  | agggacccca  | 5100 |
| ggagtgcctg  | gtccgtatct  | acattgtccg  | catagggaag | ctgcagccca  | aggaccccaa  | 5160 |
| tggaaagtgt  | gatccttaca  | tcaagatctc  | atgttgaaag | aaatcagtga  | gtgaccagga  | 5220 |
| taactacatc  | ccctgcacgc  | tggagcccgt  | tctctatgac | atgttcgagc  | tgacctgcac  | 5280 |
| tctgcctctg  | gagaaggacc  | taaagatcac  | ggagaacagg | tatgacctcc  | tctccaagga  | 5340 |
| cgaaaagatc  | ggtgagacgg  | tcgtcgacct  | ctctggaccg | ctgctgtcca  | agtttggggc  | 5400 |
| tcgctgtgga  | ctcccacaga  | cctactgtgt  | ccagcagcat | aaccagtggc  | gggaccagct  | 5460 |
| ccgcccctcc  | cagctcctcc  | acctcttctg  | agaatattcc | agagtcaagg  | cacctgtgta  | 5520 |
| ccggacagac  | cgtgtaatgt  | ttcaggataa  | ggaggagcgt | attgaagaga  | tagaggctgg  | 5580 |
| caggatccca  | aacccacacc  | tgggcccagt  | gtcacggccc | ctggctctgc  | atgtgcttca  | 5640 |
| gcagcagggc  | ctggtcccgg  | agcacgtgga  | ggtcgacctc | ctctacagcc  | ccctgcagcc  | 5700 |
| agacatcgag  | caggggaagc  | tgcagatgtg  | gagagccaga | tttccgaagg  | ccctggggcg  | 5760 |
| gcctggacct  | cccttcaaca  | tcacccacag  | tgacctgagc | aggtttttcc  | tgcggtgtat  | 5820 |
| tatctggaat  | accagagatg  | tgatectgga  | cacaagcaaa | ctcacggggg  | agaagatgag  | 5880 |
| cgacatttat  | gtgaaagggt  | ggatgattgg  | ctttgaagaa | aggttcattt  | agacagacgt  | 5940 |
| gcattatcgt  | gtccctgggag | gtgaaggcaa  | cttcaactgg | gatgccttct  | tccccttcga  | 6000 |
| ctacctgcca  | gctgagcaag  | tctgtaccat  | tgccaagaag | atctggggaca | ggaggctgga  | 6060 |
| caagactgag  | agcaaaatcc  | cagcacgagt  | ggtgttccag |             | atgacaagtt  | 6120 |

27/68

|             |             |            |            |            |            |      |
|-------------|-------------|------------|------------|------------|------------|------|
| ctcctttgat  | gatttttctgg | gctccctgca | gctcgatctc | aaccgcatgc | ccaagccagc | 6180 |
| caagacagcc  | aagaagtgtc  | ccttggacca | gctggatgat | gctttccacc | cagaatggtt | 6240 |
| tgtgtccctt  | tttgagcaga  | aaacagttaa | gggctgggtg | ccctgtgtag | cagaagaggg | 6300 |
| tgagaagaaa  | atactggcgg  | gcaagctgga | aatgaccttg | gagattgtag | cagagagtga | 6360 |
| gcatgaggag  | cgccctgctg  | gccagggccg | ggatgagccc | aacatgaacc | ctaagcttga | 6420 |
| ggacccaagg  | cgccccgaca  | cctccttcc  | gtggtttacc | tccccataca | agaccatgaa | 6480 |
| gttcatacctg | tgccggtggt  | tccggtgggc | catcatcctc | ttcatcatcc | tcttcatcct | 6540 |
| gctgctgttc  | ctggccatct  | tcctctacgc | cctcccgaac | tatgctgcca | tgaagctggt | 6600 |
| gaagcccttc  | agctgaggac  | tctcctgccc | tgtagaagg  | gccgtggggt | ccctccagc  | 6660 |
| atgggactgg  | cctgcctcct  | ccgcccagct | cgccgagctc | ctccagacct | cctaggcctg | 6720 |
| attgtcctgc  | cagggtgggc  | agacagacag | atggaccggc | ccacactccc | agagttgcta | 6780 |
| acatggagct  | ctgagatcac  | cccacttcca | tcatttccct | ctcccccaac | ccaacgcttt | 6840 |
| tttgatcag   | ctcagacata  | tttcagtata | aaacagttgg | aaccacaaaa | aaaaaaaaaa | 6900 |
| aaaaaaaaaa  | a           |            |            |            |            | 6911 |

&lt;210&gt; 19

&lt;211&gt; 6911

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 19

|             |             |             |            |             |             |      |
|-------------|-------------|-------------|------------|-------------|-------------|------|
| tcgaccgccc  | agccaggtgc  | aaaatgccgt  | gtcattggga | gactccgcag  | ccggagcatt  | 60   |
| agattacagc  | tcgacggagc  | tcgggaaggg  | cgccgggggt | ggaagatgag  | cagaagcccc  | 120  |
| tggttctcga  | acgcccggctg | acaagcgggg  | tgagcgcagg | cggggcgggg  | acccagccta  | 180  |
| gcccactgga  | gcagccgggg  | gtggcccgtt  | cccctttaag | agcaactgct  | ctaagccagg  | 240  |
| agccagagat  | tcgagccggc  | ctcgcccagc  | cagccctctc | cagcagaggg  | acccacaagc  | 300  |
| ggcgccctcg  | ccctcccagc  | ctttccgagc  | cctcttttgc | ccctggggcg  | acggggccct  | 360  |
| acacgcgcca  | agcatgctga  | gggtcttcat  | cctctatgcc | gagaacgtcc  | acacaccga   | 420  |
| caccgacatc  | agcgtgcct   | actgctccgc  | gggtgttgca | ggggtgaaga  | agagaaccaa  | 480  |
| agtcatacaag | aacagcgtga  | accctgtatg  | gaatgaggga | tttgaatggg  | acctcaagg   | 540  |
| catccccctg  | gaccagggct  | ctgagcttca  | tgtggtgggc | aaagaccatg  | agacgatggg  | 600  |
| gaggaacagg  | ttcctggggg  | aagccaaggt  | cccactccga | gaggtcctcg  | ccacccttag  | 660  |
| tctgtccgcc  | agcttcaatg  | ccccctgtct  | ggacaccaag | aagcagccca  | caggggcctc  | 720  |
| gctggctcctg | caggtgtcct  | acacaccgct  | gcctggagct | gtgcccctgt  | tcccgcctcc  | 780  |
| tactcctctg  | gagccctccc  | cgactctgcc  | tgacctggat | gtagtggcag  | acacaggagg  | 840  |
| agaggaagac  | acagaggacc  | agggaactac  | tggagatgag | gcggagccat  | tcctggatca  | 900  |
| aagcggaggg  | ccggggggctc | ccaccacccc  | aaggaaacta | ccttcacgtc  | ctccgcccc   | 960  |
| ctaccccggg  | atcaaaaagaa | agcgaagtgc  | gcctacatct | agaaagctgc  | tgctcagacaa | 1020 |
| accgcaggat  | ttccagatca  | gggtccaggt  | gatcgagggg | cgccagctgc  | cggggggtgaa | 1080 |
| catcaagcct  | gtggtcaagg  | ttaccgctgc  | agggcagacc | aagcggacgc  | ggatccacaa  | 1140 |
| gggaaacagc  | ccactcttca  | atgagactct  | tttcttcaac | ttgtttgact  | ctcctgggga  | 1200 |
| gctgtttgat  | gagcccatct  | ttatcacggt  | ggtagactct | cgttctctca  | ggacagatgc  | 1260 |
| tctcctcggg  | gagttccgga  | tggacgtggg  | caccatttac | agagagcccc  | ggcacgccta  | 1320 |
| tctcaggaag  | tggctgctgc  | tctcagaccc  | tgatgacttc | tctgctgggg  | ccagaggcta  | 1380 |
| cctgaaaaca  | agcctttgtg  | tgtggggccc  | tggggacgaa | gcgcctctgg  | agagaaaaga  | 1440 |
| cccctctgaa  | gacaaggagg  | acattgaaag  | caacctgtct | cgcccccacag | gcgtagccct  | 1500 |
| gcgaggagcc  | cacttctgcc  | tgaaggtctt  | ccgggcccag | gacttgccgc  | agatggacga  | 1560 |
| tgccgtgatg  | gacaacgtga  | aacagatctt  | tggcttcgag | agtaacaaga  | agaacttggt  | 1620 |
| ggaccccttt  | gtggaggtca  | gctttgccc   | gaaaatgctg | tgacagcaaga | tcttggaaga  | 1680 |
| gacggccaac  | cctcagtggg  | accagaacat  | cacactgcct | gccatgtttc  | cctccatgtg  | 1740 |
| cgaaaaaatg  | aggattcgta  | tcatagactg  | ggaccgcctg | actcacaatg  | acatcgctgg  | 1800 |
| taccacctac  | ctgagtatgt  | cgaaaatctc  | tgcccctgga | ggagaaaatag | aagaggagcc  | 1860 |
| tgcaggtgct  | gtcaagcctt  | cgaaaagcctc | agacttggtg | gactacctgg  | gcttccctcc  | 1920 |
| cacttttggg  | ccctgctaca  | tcaacctcta  | tggcagtcct | agagagttca  | caggcttccc  | 1980 |
| agacccttac  | acagagctca  | acacaggcaa  | gggggaaggt | gtggcttatc  | gtggccggct  | 2040 |
| tctgctctcc  | ctggagacca  | agctgggtgga | gcacagtga  | cagaaggtgg  | aggaccttcc  | 2100 |
| tgccgatgac  | atcctccggg  | tggagaagta  | ccttaggagg | cgcaagtact  | ccctgtttgc  | 2160 |
| ggccttctac  | tcagccacca  | tgctgcagga  | tgtggatgat | gccatccagt  | ttgaggtcag  | 2220 |
| catcggaagc  | tacgggaaca  | agttcgacat  | gactggcctg | ccgctggcct  | ccaccactca  | 2280 |
| gtacagccgt  | cgagtctttg  | acgggtgcc   | ctactactac | ctaccctggg  | gtaacgtgaa  | 2340 |
| acctgtgggtg | gtgctgtcat  | cctactggga  | ggacatcagc | catagaatcg  | agactcagaa  | 2400 |
| ccagctgctt  | gggattgctg  | accggctgga  | agctggcctg | gagcaggtcc  | acctggccct  | 2460 |
| gaaggcgcag  | tgctccacgg  | aggacgtgga  | ctcgctgggt | gctcagctga  | cggatgagct  | 2520 |
| catcgacagg  | tgcagccagc  | ctctgggtga  | catccatgag | acaccctctg  | ccaccacact  | 2580 |
| ggaccagtac  | ctgtaccagc  | tgcgcaccca  | tcacctgagc | caaactactg  | aggctgccct  | 2640 |
| ggccctgaag  | ctcgccacaa  | gtgagctccc  | tgcagctctg | gagcagggcg  | aggactggct  | 2700 |
| cctgcgtctg  | cgtgccctgg  | cagaggagcc  | ccagaacagc | ctgccggaca  | tcgtcatctg  | 2760 |

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| gatgctgcag  | ggagacaagc  | gtgtggcata  | ccagcgggtg  | cccgcacc    | aagtcctctt  | 2820 |
| ctcccggcgg  | ggtgccaact  | actgtggcaa  | gaattgtggg  | aagctacaga  | caatctttct  | 2880 |
| gaaatatccg  | atggagaagg  | tgcttggcgc  | ccggatgcc   | gtgcagatac  | gggtcaagct  | 2940 |
| gtggtttggg  | ctctctgtgg  | atgagaagga  | gttcaaccag  | tttgctgagg  | ggaagctgtc  | 3000 |
| tgtctttgct  | gaaacctatg  | agaacgagac  | taagttggcc  | cttggtggga  | actggggcac  | 3060 |
| aacgggcctc  | acctaccca   | agttttctga  | cgtcacgggc  | aagatcaagc  | taccaagga   | 3120 |
| cagcttcgcg  | ccctcgccg   | gctggacctg  | ggctggagat  | tggttcgtgt  | gtccggagaa  | 3180 |
| gactctgctc  | catgacatgg  | acgccgtca   | cctgagcttc  | gtggaagagg  | tgtttgagaa  | 3240 |
| ccagaccggg  | cttcccggag  | gccagtggat  | ctacatgagt  | gacaactaca  | ccgatgtgaa  | 3300 |
| cggggagaag  | gtgcttccca  | aggatgacat  | tgagtgccca  | ctgggctgga  | agtgggaaga  | 3360 |
| tgaggaatgg  | tccacagacc  | tcaaccgggc  | tgctgatgag  | caaggctggg  | agtatagcat  | 3420 |
| caccatcccc  | ccggagcgg   | agcgaagca   | ctgggtccct  | gctgagaaga  | tgtactacac  | 3480 |
| acaccgacgg  | cggcgctggg  | tgccgctgcg  | caggagggat  | ctcagccaaa  | tggaagcact  | 3540 |
| gaaaaggcac  | aggcaggcgg  | aggcggaggg  | cgagggtgg   | gagtagcct   | ctcttttttg  | 3600 |
| ctggaagtgc  | cacctcgagt  | accgcaagac  | agatgccttc  | cgcgcgcgcg  | gctggcgcgcg | 3660 |
| tcgcatggag  | ccactggaga  | agacgggggc  | ttcagctgtg  | tttgcccttg  | agggggccct  | 3720 |
| gggcggcgtg  | atggatgaca  | agagtgaaga  | ttccatgtcc  | gtctccacct  | tgagcttcgg  | 3780 |
| tgtaacaga   | cccacgattt  | cctgcatatt  | cgactatggg  | aaccgctacc  | atctacgctg  | 3840 |
| ctacatgtac  | caggcccggg  | acctggctgc  | gatggacaag  | gactcttttt  | ctgatcccta  | 3900 |
| tgccatcgtc  | tccttcctgc  | accagagcca  | gaagacgggtg | gtggtgaaga  | acacccttaa  | 3960 |
| ccccacctgg  | gaccagacgc  | tcactcttcta | cgagatcgag  | atctttggcg  | agccggccac  | 4020 |
| agttgtgtg   | caaccgcccc  | gcattgtggt  | ggagctgtac  | gacctgaca   | cttatggtgc  | 4080 |
| agacgagttt  | atgggtcgct  | gcactctgtca | accgagtctg  | gaacggatgc  | cacggtggc   | 4140 |
| ctggttccca  | ctgacgagg   | gcagccagcc  | gtcgggggag  | ctgctggcct  | cttttgagct  | 4200 |
| catccagaga  | gagaagccgg  | ccatccacca  | tattcctggt  | tttgaggtgc  | aggagacatc  | 4260 |
| aaggatcctg  | gatgagctg   | aggacacaga  | cctgccttac  | ccaccacccc  | agaggaggc   | 4320 |
| caacatctac  | atggttcctc  | agaacatcaa  | gccagcgctc  | cagcgtaccg  | ccatcgagat  | 4380 |
| cctggcatgg  | ggcctgcgga  | acatgaagag  | ttaccagctg  | gccaacatct  | cctccccag   | 4440 |
| cctcgtggta  | gagtgtgggg  | gccagacggt  | gcagtcctgt  | gtcatcagga  | acctccggaa  | 4500 |
| gaaccccaac  | tttgacatct  | gcaccctctt  | catggaagtg  | atgctgcca   | gggaggagct  | 4560 |
| ctactgcccc  | cccataccg   | tcaaggtcat  | cgataaaccg  | cagtttgccc  | gccggcctgt  | 4620 |
| ggtgggccaag | tgtaccatcc  | gctccctgga  | gagcttcctg  | tgtgaccctt  | actcggcgga  | 4680 |
| gagtcctatcc | ccacagggtg  | gcccagacga  | tgtgagccta  | ctcagtcctg  | gggaagacgt  | 4740 |
| gctcatcgac  | attgatgaca  | aggagccctt  | catccccatc  | caggaggaag  | agttcatcga  | 4800 |
| ttggtggagc  | aaattctttg  | cctccatagg  | ggagaggga   | aagtgcggct  | cctacctgga  | 4860 |
| gaaggatttt  | gacacccctga | aggtctatga  | cacacagctg  | gagaatgtgg  | aggcctttga  | 4920 |
| gggcctgtct  | gacttttgta  | acaccttcaa  | gctgtaccgg  | ggcaagacgc  | aggaggagac  | 4980 |
| agaagatcca  | tctgtgattg  | gtgaatttaa  | gggcctcttc  | aaaattttatc | ccctcccaga  | 5040 |
| agaccagcc   | atccccatgc  | ccccaaagaca | gttccaccag  | ctggccgcgc  | agggaaccca  | 5100 |
| ggagtgcctg  | gtccgtatct  | acattgtccg  | agcatttggc  | ctgcagccca  | aggaccccaa  | 5160 |
| tggaagtgct  | gatccttaca  | tcaagatctc  | catagggaag  | aaatcagtga  | gtgaccagga  | 5220 |
| taactacatc  | ccctgcacgc  | tggagcccg   | atgttggaag  | atgttcgagc  | tgacctgcac  | 5280 |
| tctgcctctg  | gagaaggacc  | taaagatcac  | tctctatgac  | tatgacctcc  | tctccaagga  | 5340 |
| cgaaaagatc  | ggtgagacgg  | tcgtcgacct  | ggagaacagg  | ctgctgtcca  | agtttggggc  | 5400 |
| tcgctgtgga  | ctcccacaga  | cctactgtgt  | ctctggaccg  | aaccagtggc  | gggaccagct  | 5460 |
| ccgcccctcc  | cagctcctcc  | acctcttctc  | ccagcagcat  | agagtcaagg  | cacctgtgta  | 5520 |
| ccggacagac  | cgtgtaatgt  | ttcaggataa  | agaatattcc  | attgaagaga  | tagaggctgg  | 5580 |
| caggatccca  | aaccacacac  | tgggcccagt  | ggaggagcgt  | ctggctctgc  | atgtgcttca  | 5640 |
| gcagcagggc  | ctgggtcccg  | agcacgtgga  | gtcacggccc  | ctctacagcc  | ccctgcagcc  | 5700 |
| agacatcgag  | cagggggaag  | tgcatgtgtg  | ggtcgacct   | tttccgaagg  | ccctggggcg  | 5760 |
| ccttggaact  | cccttcaaca  | tcacccacg   | gagagccaga  | aggtttttcc  | tgctgtgtat  | 5820 |
| tatctggaat  | accagagatg  | tgatcctgga  | tgacctgagc  | ctcacggggg  | agaagatgag  | 5880 |
| cgacatttat  | gtgaaagggt  | ggatgattgg  | ctttgaagaa  | cacaagcaaa  | agacagacgt  | 5940 |
| gcattatcgt  | tccttgggag  | gtgaaggcaa  | cttcaactgg  | aggttcattt  | tccccttcga  | 6000 |
| ctacctgcca  | gctgagcaag  | tctgtacct   | tgccaagaag  | gatgccttct  | ggaggctgga  | 6060 |
| caagactgag  | agcaaaatcc  | cagcacgagt  | ggtgttccag  | atctgggaca  | atgacaagtt  | 6120 |
| ctcctttgat  | gattttctgg  | gctccctgca  | gctcgatctc  | aaccgcatgc  | ccaagccagc  | 6180 |
| caagacagcc  | aagaagtgt   | ccttggacca  | gctggatgat  | gctttccacc  | cagaatggtt  | 6240 |
| tgtgtccctt  | tttgagcaga  | aaacagtga   | gggctggtgg  | cctgtgttag  | cagaagaggg  | 6300 |
| tgagaagaaa  | atatggcgg   | gcaagctgga  | aatgaccttg  | gagattgtag  | cagagagtga  | 6360 |
| gcatgaggag  | cggcctgctg  | gccagggccg  | ggatgagccc  | aacatgaacc  | ctaagcttga  | 6420 |
| ggacccaagg  | cgcgccgaca  | cctccttcc   | gtggtttacc  | tccccatata  | agaccatgaa  | 6480 |
| gttcatectg  | tggcggtgtt  | tccggtgggc  | catcatcctc  | ttcatcatcc  | tcttcacctc  | 6540 |
| gctgctgttc  | ctggccatct  | tcactctacg  | cttcccgaac  | tatgctgcca  | tgaagctggt  | 6600 |
| gaagcccttc  | agctgaggac  | tctcctgccc  | tgtagaaggg  | gocgtgggg   | ccccctccagc | 6660 |
| atgggactgg  | cctgcctcct  | ccgcccagct  | cggcgagctc  | ctccagacct  | cctaggcctg  | 6720 |
| attgtcctgc  | cagggtgggc  | agacagacag  | atggaccggc  | ccacactccc  | agagttgcta  | 6780 |

29/68

|            |            |             |             |            |            |      |
|------------|------------|-------------|-------------|------------|------------|------|
| acatggagct | ctgagatcac | cccactttcca | tcattttcett | ctcccccaac | ccaacgcttt | 6840 |
| tttggatcag | ctcagacata | tttcaggtata | aaacagttgg  | aaccacaaaa | aaaaaaaaaa | 6900 |
| aaaaaaaaaa | a          |             |             |            |            | 6911 |

<210> 20  
 <211> 6911  
 <212> DNA  
 <213> Homo sapiens

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| <400> 20    |             |             |             |             |             |      |
| tcgaccgccc  | agccaggtgc  | aaaatgccgt  | gtcattggga  | gactccgcag  | ccggagcatt  | 60   |
| agattacagc  | tcgacggagc  | tcgggaaggc  | cggcgggggg  | ggaagatgag  | cagaagcccc  | 120  |
| tggttctcga  | acgcccggctg | acaagcgggg  | tgagcgcagg  | cggggcgggg  | accagccta   | 180  |
| gcccactgga  | gcagccgggg  | gtggcccgtt  | cccctttaag  | agcaactgct  | ctaagccagg  | 240  |
| agccagagat  | tcgagccggc  | ctcgcgccagc | cagccctctc  | cagcgagggg  | accacaagc   | 300  |
| ggcgccctcg  | ccctcccgc   | ctttccgagc  | cctctttgcg  | ccctgggcgc  | acggggccct  | 360  |
| acacgcgcca  | agcatgctga  | gggtcttcat  | cctctatgcc  | gagaacgtcc  | acacaccgga  | 420  |
| caccgacatc  | agcgatgcct  | actgctccgc  | gggtgttgca  | gggggtgaaga | agagaaccaa  | 480  |
| agtcatacaag | aacagcgtga  | accctgtatg  | gaatgagggg  | tttgaatggg  | acctcaaggg  | 540  |
| catccccctg  | gaccagggct  | ctgagcttca  | tgtggtggct  | aaagaccatg  | agacgatggg  | 600  |
| gaggaacagg  | ttcctggggg  | aagccaaggt  | cccactccga  | gaggtcctcg  | ccacccttag  | 660  |
| tctgtccgcc  | agcttcaatg  | ccccctgct   | ggacaccaag  | aagcagccca  | caggggcctc  | 720  |
| gctggtcctg  | caggtgtcct  | acacaccgct  | gcctggagct  | gtgcccctgt  | tcccgccccc  | 780  |
| tactcctctg  | gagccctccc  | cgactctgcc  | tgacctggat  | gtagtggcag  | acacaggagg  | 840  |
| agaggaagac  | acagaggacc  | agggactcac  | tggagatgag  | gcggagccat  | tcctggatca  | 900  |
| aagcggaggc  | ccgggggctc  | ccaccacccc  | aaggaaacta  | ccttcacgtc  | ctccgccccca | 960  |
| ctaccccggg  | atcaaaaaga  | agcgaagtgc  | gcctacatct  | agaaagctgc  | tgctcagacaa | 1020 |
| accgcaggat  | ttccagatca  | gggtccaggt  | gatcgagggg  | cgccagctgc  | cgggggtgaa  | 1080 |
| catcaagcct  | gtggtcaagg  | ttaccgctgc  | agggcagacc  | aagcggacgc  | ggatccacaa  | 1140 |
| gggaaacagc  | ccactcttca  | atgagactct  | tttcttcaac  | ttgtttgact  | ctcctgggga  | 1200 |
| gctgtttgat  | gagcccatct  | ttatcacggg  | ggtagactct  | cgttctctca  | ggacagatgc  | 1260 |
| tctcctcggg  | gagttccgga  | tggacgtggg  | caccattttac | agagagcccc  | ggcacgccta  | 1320 |
| tctcaggaag  | tggctgctgc  | tctcagaccc  | tgatgacttc  | tctgctgggg  | ccagaggcta  | 1380 |
| cctgaaaaca  | agcctttgtg  | tgtctggggc  | tggggacgaa  | gcgcctctgg  | agagaaaaga  | 1440 |
| cccctctgaa  | gacaaggagg  | acattgaaag  | caacctgtct  | cggcccacag  | cggtagccct  | 1500 |
| gcgaggagcc  | cactcttgcc  | tgaaggctct  | ccgggcccag  | gacttgccgc  | agatggacga  | 1560 |
| tgccgtgatg  | gacaacgtga  | aacagatctt  | tggcttcgag  | agtaacaaga  | agaacttggt  | 1620 |
| ggaccccttt  | gtggagggtca | gctttgcggg  | gaaaatgctg  | tgacgaaga   | tcttgagaa   | 1680 |
| gacggccaac  | cctcagtggg  | accagaacat  | cacactgcct  | gccatgtttc  | cctccatgtg  | 1740 |
| cgaaaaaatg  | aggattcgta  | tcatagactg  | ggaccgcctg  | actcacaatg  | acatcgctgg  | 1800 |
| taccacctac  | ctgagtatgt  | cgaaaatctc  | tgcccctgga  | ggagaaatag  | aagaggagcc  | 1860 |
| tgacaggtgct | gtcaagcctt  | cgaaagcctc  | agacttggtg  | gactacctgg  | gcttccctcc  | 1920 |
| cactttttggg | ccctgctaca  | tcaacctcta  | tggcagctcc  | agagagttca  | caggcttccc  | 1980 |
| agacccctac  | acagagctca  | acacaggcaa  | gggggaaggt  | gtggcttatc  | gtggccggct  | 2040 |
| tctgctctcc  | ctggagacca  | agctggtgga  | gcacagtga   | cagaagggtg  | aggaccttcc  | 2100 |
| tgccgatgac  | atcctccggg  | tggagaagta  | ccttaggagg  | cgcaagtact  | ccctgtttgc  | 2160 |
| ggccttctac  | tcagccacca  | tgctgcagga  | tgtggatgat  | gccatccagt  | ttgaggtcag  | 2220 |
| catcggaac   | tacgggaaca  | agttcgacat  | gacctgcctg  | ccgctggcct  | ccaccactca  | 2280 |
| gtacagccgt  | gcagtctttg  | acgggtgcca  | ctactactac  | ctaccctggg  | gtaacgtgaa  | 2340 |
| acctgtggtg  | gtgctgtcat  | cctactggga  | ggacatcagc  | catagaatcg  | agactcagaa  | 2400 |
| ccagctgctt  | gggattgctg  | accggtgga   | agctggcctg  | gagcaggtcc  | acctggccct  | 2460 |
| gaaggcgag   | tgctccacgg  | aggacgtgga  | ctcgtgggtg  | gctcagctga  | cggatgagct  | 2520 |
| catcgaggc   | tgacgccagc  | ctctgggtga  | catccatgag  | acaccctctg  | ccaccacctc  | 2580 |
| ggaccagtac  | ctgtaccagc  | tgcgcaccca  | tcacctgagc  | caaatactg   | aggctgccct  | 2640 |
| ggccctgaag  | ctcgccca    | gtgagctccc  | tgacgtctg   | gagcaggcgg  | aggactggct  | 2700 |
| cctgcgtctg  | cgtgccctgg  | cagaggagcc  | ccagaacagc  | ctgccggaca  | tcgtcatctg  | 2760 |
| gatgctgcag  | ggagacaagc  | gtgtggcata  | ccagcgggtg  | cccggccacc  | aagtcctctt  | 2820 |
| ctcccggcgg  | gggtgccaa   | actgtggcaa  | gaattgtggg  | aagctacaga  | caatctttct  | 2880 |
| gaaatatccg  | atggagaagg  | tgccctggcg  | ccggatgcca  | gtgcagatac  | gggtcaagct  | 2940 |
| gtggtttggg  | ctctctgtgg  | atgagaagga  | gttcaaccag  | tttgctgagg  | ggaagctgtc  | 3000 |
| tgtctttgct  | gaaacctatg  | agaacgagac  | taagttggcc  | cttggtggga  | actggggcac  | 3060 |
| aacgggcctc  | acctacccca  | agttttctga  | cgtcacgggc  | aagatcaagc  | tacccaagga  | 3120 |
| cagcttccgc  | ccctcgcccg  | gctggacctg  | ggctggagat  | tggttcgtgt  | gtccggagaa  | 3180 |
| gactctgctc  | catgacatgg  | acgcgggtca  | cctgagcttc  | gtgggaagag  | tgtttgagaa  | 3240 |
| ccagaccggg  | cttcccgagg  | gccagtggtg  | ctacatgagt  | gacaactaca  | ccgatgtgaa  | 3300 |
| cgggggagaag | gtgcttccca  | aggatgacat  | tgagtgcaca  | ctgggctgga  | agtgggaaga  | 3360 |
| tgaggaatgg  | tccacagacc  | tcaaccgggc  | tgctgatgag  | caaggctggg  | agtatagcat  | 3420 |

30/68

|             |             |             |             |            |            |      |
|-------------|-------------|-------------|-------------|------------|------------|------|
| caccatcccc  | ccggagcgga  | agccgaagca  | ctgggtccct  | gctgagaaga | tgtactacac | 3480 |
| acaccgacgg  | cggcgctggg  | tgcgcctgcg  | caggagggat  | ctcagccaaa | tggaagcact | 3540 |
| gaaaaggcac  | aggcaggcgg  | aggcggaggg  | cgagggctgg  | gagtacgcct | ctcttttttg | 3600 |
| ctggaagtte  | cacctcgagt  | accgcaagac  | agatgccttc  | cgccgcgcgc | gctggcgccg | 3660 |
| tcgcatggag  | ccactggaga  | agacggggcc  | tgcagctgtg  | tttgcccttg | agggggccct | 3720 |
| gggcggcggtg | atggatgaca  | agagtgaaga  | ttccatgtcc  | gtctccacct | tgagcttcgg | 3780 |
| tgtgaacaga  | cccacgattt  | cctgcatatt  | cgactatggg  | aaccgctacc | atctacgctg | 3840 |
| ctacatgtac  | caggcccggg  | acctggctgc  | gatggacaag  | gactcttttt | ctgatcccta | 3900 |
| tgccatcgte  | tccttcctgc  | accagagcca  | gaagacgggtg | gtggtgaaga | acacccttaa | 3960 |
| ccccacctgg  | gaccagacgc  | tcatcttcta  | cgagatcgag  | atctttggcg | agccggccac | 4020 |
| agttgctgag  | caaccgcccc  | gcattgtggt  | ggagctgtac  | gaccatgaca | cttatggtgc | 4080 |
| agacgagttt  | atgggtcgct  | gcatctgtca  | accgagtctg  | gaacggatgc | cacggctggc | 4140 |
| ctgggtccca  | ctgacgaggg  | gcagccagcc  | gtcgggggag  | ctgctggcct | cttttgagct | 4200 |
| catccagaga  | gagaagccgg  | ccatccacca  | tattcctggt  | tttgaggtgc | aggagacatc | 4260 |
| aaggatcctg  | gatgagtctg  | aggacacaga  | cctgcccctac | ccaccacccc | agggagaggc | 4320 |
| caacatctac  | atggttcctc  | agaacatcaa  | cgacgcgtc   | cagcgtaccg | ccatcgagat | 4380 |
| cctggcatgg  | ggcctgcgga  | acatgaagag  | ttaccagctg  | gccaacatct | cctccccag  | 4440 |
| cctcgtggta  | gagtgtgggg  | gccagacggt  | gcagtcctgt  | gtcatcagga | acctccggaa | 4500 |
| gaaccccaac  | tttgacatct  | gcaccctctt  | catggaagtg  | atgctgccc  | gggaggagct | 4560 |
| ctactgcccc  | cccatacccg  | tcaaggtcat  | cgataaccgc  | cagtttggcc | ggcgccctgt | 4620 |
| ggtggggccag | gttaccatcc  | gctccctgga  | gagcttcctg  | tgtgaccctt | actcggcgga | 4680 |
| gagtcctacc  | ccacagggtg  | gcccagacga  | tgtgagccta  | ctcagtcctg | gggaagacgt | 4740 |
| gctcatcgac  | attgatgaca  | aggagcccct  | catccccatc  | caggaggaag | agttcatcga | 4800 |
| ttggtggagc  | aaattctttg  | cctccatagg  | ggagagggaa  | aagtgcggct | cctacctgga | 4860 |
| gaaggatttt  | gacaccctga  | aggtctatga  | cacacagctg  | gagaatgtgg | aggcctttga | 4920 |
| gggcctgtct  | gacttttgta  | acaccttcaa  | gctgtaccgg  | ggcaagacgc | aggaggagac | 4980 |
| agaagatcca  | tctgtgattg  | gtgaatttaa  | gggcctcttc  | aaaatttatc | ccctcccaga | 5040 |
| agaccagccc  | atccccatgc  | ccccaaagaca | gttccaccag  | ctggccgccc | agggacccca | 5100 |
| ggagtgcctg  | gtccgtatct  | acattgtccg  | agcattttggc | ctgcagccca | aggaccccaa | 5160 |
| tggaaagtgt  | gatccttaca  | tcaagatctc  | catagggaag  | aaatcagtg  | gtgaccagga | 5220 |
| taactacatc  | ccctgcacgc  | tggagcccgt  | atttggaag   | atgttcgagc | tgacctgcac | 5280 |
| tctgcctctg  | gagaaggacc  | taaagatcac  | tctctatgac  | tatgacctcc | tctccaagga | 5340 |
| cgaaaagatc  | ggtgagacgg  | tcgtcgacct  | ggagaacagg  | ctgctgtcca | agtttggggc | 5400 |
| tcgctgtgga  | ctcccacaga  | cctactgtgt  | ctctggaccg  | aaccagtggc | gggaccagct | 5460 |
| ccgccccctc  | cagtcctctc  | acctcttctg  | ccagcagcat  | agagtcaagg | cacctgtgta | 5520 |
| ccggacagac  | cggtgtaagt  | ttcaggataa  | agaatattcc  | attgaagaga | tagaggctgg | 5580 |
| caggatccca  | aaccacaccc  | tgggcccagt  | ggaggagcgt  | ctggctctgc | atgtgcttca | 5640 |
| gcagcagggc  | ctggctcccg  | agcacgtgga  | gtcacggccc  | ctctacagcc | ccctgcagcc | 5700 |
| agacatcgag  | caggggaagc  | tgcagatgtg  | ggtcgacct   | tttccgaagg | ccctggggcg | 5760 |
| gcctggacct  | cccttcaaca  | tcaccccacg  | gcagagccaga | aggtttttcc | tgcgttgtat | 5820 |
| tatctggaat  | accagagatg  | tgatcctgga  | tgacctgagc  | ctcacggggg | agaagatgag | 5880 |
| cgacattttat | gtgaaagggt  | ggatgattgg  | ctttgaagaa  | cacaagcaaa | agacagacgt | 5940 |
| gcgttatcgt  | tccctgggag  | gtgaaggcaa  | cttcaactgg  | aggttcattt | tccccttcga | 6000 |
| ctacctgcca  | gctgagcaag  | tctgtacct   | tgccaagaag  | gatgccttct | ggaggctgga | 6060 |
| caagactgag  | agcaaaaatcc | cagcacgagt  | ggtgttccag  | atctgggaca | atgacaagtt | 6120 |
| ctcctttgat  | gattttctgg  | gctccctgca  | gctcgatctc  | aaccgcatgc | ccaagccagc | 6180 |
| caagacagcc  | aagaagtgt   | ccttggaaca  | gctggatgat  | gctttccacc | cagaatggtt | 6240 |
| tgtgtccctt  | tttgagcaga  | aaacagtga   | gggctggtgg  | ccctgtgtag | cagaagaggg | 6300 |
| tgagaagaaa  | atactggcgg  | gcaagctgga  | aatgaccttg  | gagattgtag | cagagagtga | 6360 |
| gcatgaggag  | cggcctgctg  | gccagggccg  | ggatgagccc  | aacatgaacc | ctaagcttga | 6420 |
| ggacccaagg  | cgccccgaca  | cctccttctc  | gtggtttacc  | tccccataca | agaccatgaa | 6480 |
| gttcatectg  | tggcggcggt  | tccgggtggc  | catcatcctc  | ttcatcatcc | tcttcatcct | 6540 |
| gctgctgttc  | ctggccatct  | tcatctacgc  | cttcccgaac  | tatgctgcca | tgaagctggg | 6600 |
| gaagcccttc  | agctgaggac  | tctcctgccc  | tgtagaaggg  | gccgtggggg | ccctccagc  | 6660 |
| atgggactgg  | cctgcctcct  | ccgcccagct  | cggcgagctc  | ctccagacct | cctaggcctg | 6720 |
| attgtcctgc  | caggggtggc  | agacagacag  | atggaccggc  | ccacactccc | agagttgcta | 6780 |
| acatggagct  | ctgagatcac  | cccacttcca  | tatttctctt  | ctcccccaac | ccaacgcttt | 6840 |
| tttgatcag   | ctcagacata  | tttcagtata  | aaacagttag  | aaccacaaaa | aaaaaaaaaa | 6900 |
| aaaaaaaaaa  | a           |             |             |            |            | 6911 |

&lt;210&gt; 21

&lt;211&gt; 6909

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 21

tcgaccgccc agccaggtgc aaaatgccgt gtcattggga gactccgcag ccggagcatt

60

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| agattacagc  | tcgacggagc  | tcgggaagg   | cgccgggggt  | ggaagatgag  | cagaagcccc  | 120  |
| tgttctcgga  | acgccggctg  | acaagcggg   | tgagcgcagg  | cggggcgggg  | acccagccta  | 180  |
| gcccactgga  | gcagccgggg  | gtggcccgtt  | cccctttaag  | agcaactgct  | ctaagccagg  | 240  |
| agccagagat  | tcgagccggc  | ctcgcccagc  | cagccctctc  | cagcgagggg  | acccacaagc  | 300  |
| ggcgccctcg  | ccctcccagc  | ctttccgagc  | cctctttgcg  | ccctgggcgc  | acggggccct  | 360  |
| acacgcgcca  | agcatgctga  | gggtcttcat  | cctctatgcc  | gagaacgtcc  | acacacccga  | 420  |
| caccgacatc  | agcgatgcct  | actgctccgc  | gggtgttgca  | ggggtgaaga  | agagaaccaa  | 480  |
| agtcatacaag | aacagcgtga  | accctgtatg  | gaatgaggga  | tttgaatggg  | acctcaaggg  | 540  |
| catccccctg  | gaccagggct  | ctgagcttca  | tgtgggtggtc | aaagaccatg  | agacgatggg  | 600  |
| gaggaacagg  | ttcctggggg  | aagccaaggt  | cccactccga  | gaggtcctcg  | ccacccctag  | 660  |
| tctgtccgcc  | agcttcaatg  | ccccctgct   | ggacaccaag  | aagcagccca  | caggggcctc  | 720  |
| gctggtcctg  | caggtgtcct  | acacaccgct  | gcctggagct  | gtgcccctgt  | tcccccccc   | 780  |
| tactcctctg  | gagccctccc  | cgactctgcc  | tgacctggat  | gtagtggcag  | acacaggagg  | 840  |
| agaggaagac  | acagaggacc  | agggactcac  | tggagatgag  | gcgagaccat  | tcttggtatca | 900  |
| aagcgggaggc | ccgggggctc  | ccaccacccc  | aaggaaacta  | ccttcacgtc  | ctccgcccc   | 960  |
| ctaccccggg  | atcaaaaagaa | agcgaagtgc  | gcctacatct  | agaaagctgc  | tgtcagacaa  | 1020 |
| accgcaggat  | ttccagatca  | gggtccaggt  | gatcgagggg  | cgccagctgc  | cgggggtgaa  | 1080 |
| catcaagcct  | gtggtcaagg  | ttaccgctgc  | agggcagacc  | aagcggacgc  | ggtaccacaa  | 1140 |
| gggaaacagc  | ccactcttca  | atgagactct  | tttcttcaac  | ttgtttgact  | ctcctgggga  | 1200 |
| gctgtttgat  | gagcccactc  | ttatcacggg  | ggtagactct  | cgttctctca  | ggacagatgc  | 1260 |
| tctcctcggg  | gagttccgga  | tggacgtggg  | caccatttac  | agagagcccc  | ggcacgccta  | 1320 |
| tctcaggaag  | tggctgctgc  | tctcagaccc  | tgatgacttc  | tctgctgggg  | ccagaggcta  | 1380 |
| cctgaaaaca  | agcctttgtg  | tgctggggcc  | tggggacgaa  | gcgcctctgg  | agagaaaaga  | 1440 |
| cccctctgaa  | gacaaggagg  | acattgaaag  | caacctgctc  | cggcccacag  | gcgtagccct  | 1500 |
| gcgaggagcc  | cacttctgcc  | tgaaggtcct  | cggggccgag  | gacttgccgc  | agatggacga  | 1560 |
| tgcggtgatg  | gacaacgtga  | aacagatctt  | tggcttcgag  | agtaacaaga  | agaacttggg  | 1620 |
| ggaccccttt  | gtggaggtca  | gctttgcggg  | gaaaatgctg  | tgcagcaaga  | tcttgagaaa  | 1680 |
| gacggccaac  | cctcagtggg  | accagaacat  | cacactgcct  | gccatgtttc  | cctccatgtg  | 1740 |
| cgaaaaaatg  | aggattcgta  | tcatagactg  | ggaccgcctg  | actcacaatg  | acatcgtggc  | 1800 |
| taccacctac  | ctgagtatgt  | cgaaaatctc  | tgcccctgga  | ggagaaaatg  | aagaggagcc  | 1860 |
| tgcaggtgct  | gtcaagcctc  | cgaaagcctc  | agacttggat  | gactacctgg  | gcttccctcc  | 1920 |
| cacttttggg  | ccctgctaca  | tcaacctcta  | tggcagtccc  | agagagttca  | caggcttccc  | 1980 |
| agacccttac  | acagagctca  | acacaggcaa  | gggggaagg   | gtggccttatc | gtggccggct  | 2040 |
| tctgctctcc  | ctggagacca  | agctggtgga  | gcacagtga   | cagaaggtgg  | aggaccttcc  | 2100 |
| tgcggtatgac | atcctccggg  | tggagaagta  | ccttaggagg  | cgcaagtact  | ccctgtttgc  | 2160 |
| ggccttctac  | tcagccacca  | tgtgcagga   | tgtggatgat  | gccatccagt  | ttgaggtcag  | 2220 |
| catcggaac   | tacgggaaca  | agttcgacat  | gacctgcctg  | ccgctggcct  | ccaccactca  | 2280 |
| gtacagccgt  | gcagtctttg  | acgggtgcca  | ctactactac  | ctaccctggg  | gtaacgtgaa  | 2340 |
| acctgtgggt  | gtgctgtcat  | cctactggga  | ggacatcagc  | catagaatcg  | agactcagaa  | 2400 |
| ccagctgctt  | gggattgctg  | accggctgga  | agctggcctg  | gagcaggtcc  | acctggccct  | 2460 |
| gaaggcgcag  | tgctccacgg  | aggacgtgga  | ctcgctgggtg | gctcagctga  | cggatgagct  | 2520 |
| catcgagggc  | tgcagccagc  | ctctgggtga  | catccatgag  | acaccctctg  | ccaccacact  | 2580 |
| ggaccagtac  | ctgtaccagc  | tgcgcaccca  | tcaactgagc  | caaatactg   | aggctgcctt  | 2640 |
| ggccctgaag  | ctcgccaca   | gtgagctccc  | tgcagctctg  | gagcaggcgg  | aggactggct  | 2700 |
| cctgcgtctg  | cgtgccctgg  | cagaggagcc  | ccagaacagc  | ctgcccggaca | tcgtcatctg  | 2760 |
| gatgctgcag  | ggagacaagc  | gtgtggcata  | ccagcgggtg  | cccgccacc   | aagtcctctt  | 2820 |
| ctcccgcgcg  | ggtgccaaact | actgtggcaa  | gaattgtggg  | aagctacaga  | caatctttct  | 2880 |
| gaaatatccg  | atggagaagg  | tgcctggcgc  | ccggatgcc   | gtgcagatac  | gggtcaagct  | 2940 |
| gtggtttggg  | ctctctgtgg  | atgagaagga  | gttcaaccag  | tttgctgagg  | ggaagctgtc  | 3000 |
| tgtctttgct  | gaaacctatg  | agaacgagac  | taagttggcc  | cttgttggga  | actggggcac  | 3060 |
| aacgggcctc  | acctacccca  | agttttctga  | cgtcacgggc  | aagatcaagc  | tacccaagga  | 3120 |
| cagcttccgc  | ccctcggccg  | gctggacctg  | ggctggagat  | tgggtcctgt  | gtccggagaa  | 3180 |
| gactctgctc  | catgacatgg  | acgccgggtc  | cctgagcttc  | gtggaagagg  | tgtttgagaa  | 3240 |
| ccagacccgg  | cttcccggag  | gccagtggat  | ctacatgagt  | gacaactaca  | ccgatgtgaa  | 3300 |
| cggggagaag  | gtgcttccca  | aggatgacat  | tgagtggcca  | ctgggctgga  | agtgggaaga  | 3360 |
| tgaggaatgg  | tccacagacc  | tcaaccgggc  | tgtcgatgag  | caaggctggg  | agtatagcat  | 3420 |
| caccatcccc  | ccggagcgga  | agccgaagca  | ctgggtccct  | gctgagaaga  | tgtactacac  | 3480 |
| acaccgacgg  | cggcgtggtg  | tgcgcctggc  | caggagggat  | ctcagccaaa  | tggaaagcact | 3540 |
| gaaaaggcac  | aggcaggcgg  | aggcggaggg  | cgagggtggg  | gagtacgcct  | ctcttttggg  | 3600 |
| ctgggaagttc | cacctcgagt  | accgcaaggc  | agatgccttc  | cgccgcggcc  | gctggcgccg  | 3660 |
| tcgcatggag  | ccactggaga  | agacggggcc  | tgcagctgtg  | tttgcccttg  | agggggccct  | 3720 |
| gggcggcggtg | atggatgaca  | agagtgaaga  | ttccatgtcc  | gtctccacct  | tgagcttcgg  | 3780 |
| tgtgaacaga  | cccacgattt  | cctgcatatt  | cgactatggg  | aaccgctacc  | atctacgtctg | 3840 |
| ctacatgtac  | caggcccggg  | acctggctgc  | gactggacaag | gactcttttt  | ctgatcccta  | 3900 |
| tgccatcgtc  | tccttcctgc  | accagagcca  | gaagacgggtg | gtggtgaaga  | acacccttaa  | 3960 |
| ccccacctgg  | gaccagacgc  | tcattcttcta | cgagatcgag  | atctttggcg  | agccggccac  | 4020 |
| agttgctgag  | caaccgcccc  | gcattgtggg  | ggagctgtac  | gaccatgaca  | cttatggtgc  | 4080 |

32/68

|            |             |             |            |             |             |      |
|------------|-------------|-------------|------------|-------------|-------------|------|
| agacgagttt | atggggtcgct | gcacgtgtca  | accgagtcgt | gaacggatgc  | cacggctggc  | 4140 |
| ctgggtccca | ctgacgaggg  | gcagccagcc  | gtcgggggag | ctgctggcct  | cttttgagct  | 4200 |
| catccagaga | gagaagccgg  | ccatccacca  | tattcctggg | tttgaggtgc  | aggagacatc  | 4260 |
| aaggatcctg | gatgagtcct  | aggacacaga  | cctgccctac | ccaccacccc  | agagggaggg  | 4320 |
| caacatctac | atggttcctc  | agaacatcaa  | gccagcgctc | cagcgtagcg  | ccatcgagat  | 4380 |
| cctgcgtggg | ggcctgcgga  | acatgaagag  | ttaccagctg | gccaacatct  | cctccccag   | 4440 |
| cctcgtggga | gagtggtggg  | gccagacggg  | gcagtcctgt | gtcatcagga  | acctccggaa  | 4500 |
| gaaccccaac | tttgacatct  | gcaccctctt  | catggaagtg | atgctgcccc  | gggaggagct  | 4560 |
| ctactgcccc | cccatcaccg  | tcaaggtcat  | cgataaccgc | cagtttgccc  | gccggcctgt  | 4620 |
| ggtggggccg | tgtaccatcc  | gctccctgga  | gagcttcctg | tgtgaccctt  | actcggcgga  | 4680 |
| gagtcctatc | ccacaggggt  | gcccagacga  | tgtgagccta | ctcagtcctg  | gggaagacgt  | 4740 |
| gctcatcgac | attgatgaca  | aggagcccc   | catccccatc | caggagggaag | agttcatcga  | 4800 |
| ttggtggagc | aaattctttg  | cctccatagg  | ggagagggaa | aagtgcgggt  | cctacctgga  | 4860 |
| gaaggatttt | gacaccctga  | aggctctatg  | cacacagctg | gagaatgtgg  | aggcctttga  | 4920 |
| gggcctgtct | gactttttgt  | acaccttcaa  | gctgtaccgg | ggcaagacgc  | aggaggagac  | 4980 |
| agaagatcca | tctgtgattg  | gtgaatttaa  | gggcctcttc | aaaattttat  | ccctcccaga  | 5040 |
| agaccagccc | atccccatgc  | ccccaaagaa  | gttccaccag | ctggccgccc  | agggacccca  | 5100 |
| ggagtgcctg | gtccgtatct  | acattgtccg  | agcatttggt | ctgcagccca  | aggaccccaa  | 5160 |
| tggaaagtgt | gatccttaca  | tcaagatctc  | catagggaag | aaatcagtga  | gtgaccagga  | 5220 |
| taactacatc | ccctgcacgc  | tggagccgtg  | atttggaag  | atgttcgagc  | tgacctgcac  | 5280 |
| tctgcctctg | gagaaggacc  | taaagatcac  | tctctatgac | tatgacctcc  | tctccaagga  | 5340 |
| cgaaaagatc | ggtgagacgg  | tcgtcgacct  | ggagaacagg | ctgctgtcca  | agtttggggc  | 5400 |
| tcgctgtgga | ctcccacaga  | cctactgtgt  | ctctggaccg | aaccagtggc  | gggaccagct  | 5460 |
| ccgccccctc | cagctcctcc  | acctcttctg  | ccagcagcat | agagtcaagg  | cacctgtgta  | 5520 |
| ccggacagac | cgtgtaattg  | ttcaggataa  | agaattattc | attgaagaga  | tagaggctgg  | 5580 |
| caggatccca | aaccacacac  | tgggcccagt  | ggaggagcgt | ctggctctgc  | atgtgcttca  | 5640 |
| gcagcagggc | ctggtccccg  | agcacgtgga  | gtcacggccc | ctctacagcc  | ccctgcagcc  | 5700 |
| agacatcgag | cagggggaagc | tgcagatgtg  | ggtcgacctc | tttccgaagg  | ccctggggcg  | 5760 |
| gcctggacct | cccttcaaca  | tcacccacag  | gagagccaga | agggttttcc  | tgcgttgtat  | 5820 |
| tatctggaat | accagagatg  | tgatcctgga  | tgacctgagc | ctcacggggg  | agaagatgag  | 5880 |
| cgacatttat | gtgaaagggt  | ggatgattgg  | ctttgaagaa | cacaagcaaa  | agacagacgt  | 5940 |
| gcattatcgt | tccctgggag  | gtgaaggcaa  | cttcaactgg | agggttcattt | tccccctcga  | 6000 |
| ctacctgcca | gctgagcaag  | tctgtaccat  | tgccaagaag | gatgccttct  | ggaggctgga  | 6060 |
| caagactgag | caaaatccca  | gcacgagtgg  | tgttccagat | ctgggacaat  | gacaagttct  | 6120 |
| cctttgatga | ttttctgggc  | tccctgcagc  | tcgatctcaa | ccgcatgccc  | aagccagcca  | 6180 |
| agacagccaa | gaagtgcctc  | ttggaccagc  | tggatgatgc | tttccaccca  | gaatggtttg  | 6240 |
| tgtccctttt | tgagcagaaa  | acagtgaagg  | gctgggtggc | ctgtgtagca  | gaagaggggtg | 6300 |
| agaagaaaat | actggcgggc  | aagctggaaa  | tgaccttgga | gattgtagca  | gagagtgagc  | 6360 |
| atgaggagcg | gcctgctggc  | cagggccggg  | atgagcccaa | catgaaccct  | aagcttgagg  | 6420 |
| acccaaggcg | ccccacacac  | tccttcctgt  | ggtttacctc | ccatacaag   | accatgaagt  | 6480 |
| tcacacctgt | gcggcggttc  | cggtgggcca  | tcacacctct | catcatcctc  | ttcatcctgc  | 6540 |
| tgtctgttct | ggccatcttc  | atctacgcct  | tcccgaacta | tgtgcccag   | aagctgggtga | 6600 |
| agcccttcag | ctgaggactc  | tcctgcccctg | tagaaggggc | cgtgggggtcc | cctccagcat  | 6660 |
| gggactggcc | tgccctcctc  | gcccagctcg  | gcgagctcct | ccagacctcc  | taggcctgat  | 6720 |
| tgtcctgcca | gggtgggcag  | acagacagat  | ggaccggccc | acactcccag  | agttgctaac  | 6780 |
| atggagctct | gagatcacc   | cacttccatc  | atttccttct | ccccaacccc  | aacgcttttt  | 6840 |
| tggatcagct | cagacatatt  | tcagtataaa  | acagttggaa | ccacaaaaaa  | aaaaaaaaaa  | 6900 |
| aaaaaaaaaa |             |             |            |             |             | 6909 |

<210> 22  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 22  
 tgggacctca agggcatccc

20

<210> 23  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 23  
 accatgctgc aggatgtgga

20

<210> 24  
 <211> 20

33/68

|                    |            |            |            |            |            |     |
|--------------------|------------|------------|------------|------------|------------|-----|
| <212> DNA          |            |            |            |            |            |     |
| <213> Homo sapiens |            |            |            |            |            |     |
| <400> 24           |            |            |            |            |            |     |
| gggaggtgaa         | ggcaacttca |            |            |            | 20         |     |
| <210> 25           |            |            |            |            |            |     |
| <211> 20           |            |            |            |            |            |     |
| <212> DNA          |            |            |            |            |            |     |
| <213> Homo sapiens |            |            |            |            |            |     |
| <400> 25           |            |            |            |            |            |     |
| ctcacggggg         | agaagatgag |            |            |            | 20         |     |
| <210> 26           |            |            |            |            |            |     |
| <211> 20           |            |            |            |            |            |     |
| <212> DNA          |            |            |            |            |            |     |
| <213> Homo sapiens |            |            |            |            |            |     |
| <400> 26           |            |            |            |            |            |     |
| ctgtggcggc         | gtttccggtg |            |            |            | 20         |     |
| <210> 27           |            |            |            |            |            |     |
| <211> 20           |            |            |            |            |            |     |
| <212> DNA          |            |            |            |            |            |     |
| <213> Homo sapiens |            |            |            |            |            |     |
| <400> 27           |            |            |            |            |            |     |
| acatcaagga         | tcctggatga |            |            |            | 20         |     |
| <210> 28           |            |            |            |            |            |     |
| <211> 20           |            |            |            |            |            |     |
| <212> DNA          |            |            |            |            |            |     |
| <213> Homo sapiens |            |            |            |            |            |     |
| <400> 28           |            |            |            |            |            |     |
| ctgtggcggc         | gtttccggtg |            |            |            | 20         |     |
| <210> 29           |            |            |            |            |            |     |
| <211> 20           |            |            |            |            |            |     |
| <212> DNA          |            |            |            |            |            |     |
| <213> Homo sapiens |            |            |            |            |            |     |
| <400> 29           |            |            |            |            |            |     |
| acagacgtgc         | attatcgttc |            |            |            | 20         |     |
| <210> 30           |            |            |            |            |            |     |
| <211> 20           |            |            |            |            |            |     |
| <212> DNA          |            |            |            |            |            |     |
| <213> Homo sapiens |            |            |            |            |            |     |
| <400> 30           |            |            |            |            |            |     |
| aagactgaga         | gcaaaatccc |            |            |            | 20         |     |
| <210> 31           |            |            |            |            |            |     |
| <211> 507          |            |            |            |            |            |     |
| <212> DNA          |            |            |            |            |            |     |
| <213> Homo sapiens |            |            |            |            |            |     |
| <400> 31           |            |            |            |            |            |     |
| tcgaccgccc         | agccaggtgc | aaaatgccgt | gtcattggga | gactccgcag | ccggagcatt | 60  |
| agattacagc         | tcgacggagc | tcgggaaggg | cggcgggggt | ggaagatgag | cagaagcccc | 120 |
| tgttctcggg         | acgccggctg | acaagcgggg | tgagcgcagg | cggggcgggg | accagccta  | 180 |
| gcccactgga         | gcagccgggg | gtggcccgtt | cccctttaag | agcaactgct | ctaagccagg | 240 |
| agccagagat         | tcgagccggc | ctcgcccagc | cagccctctc | cagcgagggg | accacaagc  | 300 |
| ggcgcctcgg         | ccctcccgc  | ctttccgagc | cctctttgcg | ccctgggcgc | acggggccct | 360 |
| acacgcacca         | agcatgctga | gggtcttcat | cctctatgcc | gagaacgtcc | acacacccga | 420 |



34/68

caccgacatc agcgatgcct actgctccgc ggtgtttgca ggtaggaggg gccgaccacc 480  
ctcgccgggg tcgggggtggg gtagagg 507

<210> 32  
<211> 183  
<212> DNA  
<213> Homo sapiens

<400> 32  
aaaggcggga tgtgtctctc cattctccct tttgtgtctc ttgtaggggt gaagaagaga 60  
accaaagtca tcaagaacag cgtgaaccct gtatggaatg aggtatgtga gtttttctcc 120  
ttccttttct ctctgtctgc tgcagggggc ttgggaggag gtgccttctc agcagtgtcc 180  
ttg 183

<210> 33  
<211> 264  
<212> DNA  
<213> Homo sapiens

<400> 33  
cattcatgaa tgcctactca gtgccctggg ggcacgaagg tgaaccagac acagtctctt 60  
ctcctagagg gccataggtt aagatgcctt ttctcttttt cttccagggg tttgaatggg 120  
acctcaaggg catccccctg gaccagggct ctgagcttca tgtgggtggc aaagaccatg 180  
agacgatggg gaggaacagg taagggtggc agaggggggt gctccatggc ttgaagggtg 240  
aggtaggatt gtggagtata caga 264

<210> 34  
<211> 223  
<212> DNA  
<213> Homo sapiens

<400> 34  
cagaagagcc aggggtgcctt aggctagttt tctacatttg acttctctct cctctcaggt 60  
tcctggggga agccaaggct ccactccgag aggtcctcgc cacccttagt ctgtccgcca 120  
gcttcaatgc cccctgctg gacaccaaga agcagccac aggggtaagt gcccatcagc 180  
ctctgccagg ttaagggtcca aggcattgcc aggtggcttc ctc 223

<210> 35  
<211> 224  
<212> DNA  
<213> Homo sapiens

<400> 35  
cagtggctccg aggccagcgc accaacctgt ccccccacgtc tcattctcttc caggcctcgc 60  
tggtcctgca ggtgtcctac acaccgtgc ctggagctgt gccctgttc ccgcccccta 120  
ctcctctgga gccctccccg actctgcctg acctggatgt agtggcaggt gggtagccca 180  
cggtggcctg gctgggcccc agcaagaatg gccggcagtg gcac 224

<210> 36  
<211> 315  
<212> DNA  
<213> Homo sapiens

<400> 36  
aggggagagg gcagggccag agggccaggc ctcattaggg ccctctcctc ttagacacag 60  
gaggagagga agacacagag gaccagggac tcaactggaga tgaggcggag ccattcctgg 120  
atcaaagcgg agggccgggg gctcccacca ccccaaggaa actaccttca cgtcctccgc 180  
cccactaccc cgggatcaaa agaaagcgaa gtgcgcctac atctagaaaag ctgctgtcag 240  
acaaaccgca ggatttccag gtgatgaacg ggctttctct gaccccaggc tcctcttcag 300  
ccatcagctg cgggt 315

<210> 37  
<211> 249  
<212> DNA  
<213> Homo sapiens

35/68

<400> 37  
 ccagtgggtga gatgggtccct gagattttctg actcttgggg tggatgggtgg gtgggtcctta 60  
 actcttcccc cttctggctt tcagatcagg gtccagggtga tcgaggggag ccagctgccg 120  
 ggggtgaaca tcaagcctgt ggtcaagggt accgctgcag ggcagaccaa gcgagcgcg 180  
 atccacaagg gaaacagccc actcttcaat gaggtgggag acatggggca tgagggcaga 240  
 accttgtgg 249

<210> 38  
 <211> 185  
 <212> DNA  
 <213> Homo sapiens

<400> 38  
 ccctggcctg agggatcagc aggcactgat atgtctctct ttgctctgaa ccaacagact 60  
 cttttcttca acttgtttga ctctcctggg gagctgtttg atgagcccat ctttatcacg 120  
 gtatgtctca gcagtcaaag tgttctccgt gggctgtatg tatgcacata ggtgtcagtg 180  
 cacac 185

<210> 39  
 <211> 196  
 <212> DNA  
 <213> Homo sapiens

<400> 39  
 aagagctatt ggggtggccg tgtggggccac atgtccctgt gaatgtgagc catgatcttt 60  
 ctctgcaggg ggtagactct cgctctctca ggacagatgc tctcctcggg gagttccggg 120  
 taattgctta ttttctaaaa gcagtcagtt ctcacttctc cgtgttggtg gagcctctgt 180  
 ggaccatggg cagggg 196

<210> 40  
 <211> 178  
 <212> DNA  
 <213> Homo sapiens

<400> 40  
 tggaatcgta taatgcacca cactttatctt aacgctttgg cggcaagagt ttgattttgtg 60  
 tctcctctct tgattgcaga tggacgtggg caccatttac agagagcccc gtgagttctc 120  
 accactttgg ccgtatcctt gcatttttgtt tctggaggct gattggggag actcattt 178

<210> 41  
 <211> 231  
 <212> DNA  
 <213> Homo sapiens

<400> 41  
 ggggtcttct gattctggga tcaccaaagg atgttgtctc tcttagggca cgcctatctc 60  
 aggaagtggc tgctgtctct agaccctgat gacttctctg ctggggccag aggctacctg 120  
 aaaacaagcc tttgtgtgct ggggcctggg gacgaagcgc ctgtgagtag atttccttg 180  
 gtcttcctta cgggtccccc cgcggcactt ggttgcggag gcaccaaacc a 231

<210> 42  
 <211> 247  
 <212> DNA  
 <213> Homo sapiens

<400> 42  
 gtcaaaaccc tgtgtcagg agcgcagtaa ggaacgtatt tggttttctt ttagctgga 60  
 gagaaaagac cctctgaag acaaggagga cattgaaagc aacctgctcc ggcccacagg 120  
 cgtagccctg cgaggagccc acttctgcct gaaggctctc cgggccgagg acttgccgca 180  
 gagtgcgtgg ggcgcgccct tgggtgggag gtctgcagga ggctggaggc gcagggtctg 240  
 tgggggt 247

<210> 43  
 <211> 179  
 <212> DNA  
 <213> Homo sapiens

36/68

<400> 43  
 caggcagtga ctggtgtgtc cctcttccca gtggacgatg ccgtgatgga caacgtgaaa 60  
 cagatctttg gcttcgagag taacaagaag aacttggtgg acccctttgt ggaggtcagc 120  
 tttgcgggga aaatggttaag gagcaaggga gcaggagggt tctctcggga ggggacggg 179

<210> 44  
 <211> 202  
 <212> DNA  
 <213> Homo sapiens

<400> 44  
 ccccggggga gccagagtc cccatggagc tgatcaactt gtcccctccc tgtgtcttct 60  
 agctgtgcag caagatcttg gagaagacgg ccaaccctca gtggaaccag aacatcacac 120  
 tgccctgccat ggtgagcctc ctgtccccag caaacccaag gaggccctg gggctctggg 180  
 cttcgggagg tccagggtc ct 202

<210> 45  
 <211> 167  
 <212> DNA  
 <213> Homo sapiens

<400> 45  
 gggaggggct gttctatctt caaaaggact cttctcccaa cagcctcta ttccttctc 60  
 agtttccctc catgtgcgaa aaaatgagga ttcgtatcat agactgggtga gttctgagtc 120  
 ttggagtctt tagggcgggc tgtcctgagg gggcgctccc tcagttt 167

<210> 46  
 <211> 220  
 <212> DNA  
 <213> Homo sapiens

<400> 46  
 tgtggcctga gttcctttcc tgtgtcaggc cctctctgct cccttgctct ctaggggaccg 60  
 cctgactcac aatgacatcg tggctaccac ctacctgagt atgtcgaaaa tctctgcccc 120  
 tggaggagaa atagaaggta tgttccctct tcgttctgcc ctttgacccc ctgtgctctc 180  
 cccccctcta tccagcttac acttctagtt ttgagagttt 220

<210> 47  
 <211> 172  
 <212> DNA  
 <213> Homo sapiens

<400> 47  
 acagcctggt catgtaaccc gtccttctcc cagccatgcc caccctaacc ctttttccat 60  
 ttcttttacg ttcagaggag cctgcagggt ctgtcaagcc ttcgaaagcc tcagactgta 120  
 cgttgctgtc accttgggga caaccagggg agtggggcct tgggttttgg ct 172

<210> 48  
 <211> 200  
 <212> DNA  
 <213> Homo sapiens

<400> 48  
 ccgaccctc tgattgccac ttgtgtctcc cagtggatga ctacctgggc ttcctcccca 60  
 cttttgggccc ctgtacatc aacctctatg gcagtcccag agagttcaca ggcttcccag 120  
 acccctacac agagctcaac acaggcaagg taagccggct ggagccctgg caagggcagg 180  
 atgccacatg cccagggtggg 200

<210> 49  
 <211> 217  
 <212> DNA  
 <213> Homo sapiens

<400> 49  
 cctccccctc gtctccccctg ctcttctgtga cctgacctcc ctggcagggg gaaggtgtgg 60  
 cttatcgtgg ccggcttctg ctctccccctg agaccaagct ggtggagcac agtgaacaga 120  
 aggtggagga ccttccctgcg gatgacatcc tccgggtgga ggtgaggggt gtggctctgg 180

37/68

gtgggagctg ggcgtcgggg cagggaaggg atggcca 217

<210> 50  
 <211> 269  
 <212> DNA  
 <213> Homo sapiens

<400> 50  
 agcctgggtg cttttctttg ctccctccgt gaccctctgg tctactctct gctctcagaa 60  
 gtaccttagg aggcgcaagt actccctgtt tgcggccttc tactcagcca ccatgctgca 120  
 ggatgtggat gatgccatcc agtttgaggt cagcatcggg aactacggga acaagttcga 180  
 catgacctgc ctgccgctgg cctccaccac tcagtacagc cgtgcagtct ttgacggtga 240  
 ggcagtgtc ctggctggga ccccgatca 269

<210> 51  
 <211> 225  
 <212> DNA  
 <213> Homo sapiens

<400> 51  
 actcctggca cagcgtcag gccctgtctt ccattccagg gtgccactac tactacctac 60  
 cctggggtaa cgtgaaacct gtgggtgtgc tgtcacccta ctgggaggac atcagccata 120  
 gaatcgagac tcagaaccag ctgcttgga ttgctgaccg gctgggtgagt gaaaacttgc 180  
 ccaaagctgc acatgcctat gcatgcacct gctacccccg ctgca 225

<210> 52  
 <211> 227  
 <212> DNA  
 <213> Homo sapiens

<400> 52  
 ggggtccagca tgcaccctct gccctgtggt gacacacctg acccttgcct gccattcca 60  
 caggaagctg gcctggagca ggtccacctg gccctgaagg cgcagtgtc cacggaggac 120  
 gtggactcgc tgggtggtca gctgacggat gagtcctcag caggctgcag gtagggggga 180  
 cctggcgccc ctgggtgccc cctctcctgg ctcaactggg cctggtt 227

<210> 53  
 <211> 303  
 <212> DNA  
 <213> Homo sapiens

<400> 53  
 tgggagaccc tgggctcatc aggcgcattc catctgtccg tccctcacag ccagcctctg 60  
 ggtgacatcc atgagacacc ctctgccacc cacctggacc agtacctgta ccagctgcgc 120  
 acccatcacc tgagccaaat cactgaggct gccctggccc tgaagctcgg ccacagttag 180  
 ctccctgcag ctctggagca ggcgaggagc tggctcctgc gtctgcgtgc cctggcagag 240  
 gaggtaatta agcctggggg tgcctttctt cttctgctct cctgctgcct ggaacatcag 300  
 aac 303

<210> 54  
 <211> 272  
 <212> DNA  
 <213> Homo sapiens

<400> 54  
 cgtgggcctg gtgtgtcacc atccccaccc cgaccaccac cctctgttca gcccagaac 60  
 agcctgccgg acatcgtcat ctggatgctg cagggagaca agcgtgtggc ataccagcgg 120  
 gtgcccgccc accaagtcct cttctcccgg cggggtgcc actactgtgg caagaattgt 180  
 gggaagctac agacaatctt tctgaaagtg agttttctt ttccaagtca tgatcgtatt 240  
 tccaacataa ggcctttctc ccattctctg ct 272

<210> 55  
 <211> 219  
 <212> DNA  
 <213> Homo sapiens

38/68

<400> 55  
 tgtggggtttc tgctcttctt cggtagccag tatccgatgg agaaggtgcc tggcgcccgg 60  
 atgccagtgc agatacgggt caagctgtgg tttgggctct ctgtggatga gaaggagttc 120  
 aaccagtttg ctgaggggaa gctgtctgtc tttgctgaaa ccgtgagtac ctgccagccc 180  
 ccacctctgc ctccactac ctggagctgc cttggcccc 219

<210> 56  
 <211> 292  
 <212> DNA  
 <213> Homo sapiens

<400> 56  
 tgcctccac tacctggagc tgccttggcc cccttcacgc ctcattcttc ctggccctcc 60  
 agtatgagaa cgagactaag ttggcccttg ttgggaactg gggcacaacg ggcctcacct 120  
 accccaagtt ttctgacgtc acgggcaaga tcaagctacc caaggacagc ttccgcccct 180  
 cggccgggtg gacctgggct ggagattggt tcgtgtgtcc ggagaagacg tgagtcgtgg 240  
 gcagggaggg ctggggagag ccaggccagg ctgcccacca tggactgcac cc 292

<210> 57  
 <211> 242  
 <212> DNA  
 <213> Homo sapiens

<400> 57  
 tggatggggg cctctccagc agagcagcag agactctgac cagccctcct ccacagtctg 60  
 ctccatgaca tggacgccgg tcacctgagc ttcgtggaag aggtgtttga gaaccagacc 120  
 cggcttcccg gaggccagtg gatctacatg agtgacaact acaccgatgt ggtaaagcag 180  
 gcactcaggg gcaggtgggg tctagacatt tggctctctg aggcacctgg tgctcagggg 240  
 ca 242

<210> 58  
 <211> 215  
 <212> DNA  
 <213> Homo sapiens

<400> 58  
 tcacatctgt ctgtctcttc tcattgcttg cctgttcggt tttgtcctta gaacggggag 60  
 aaggtgcttc ccaaggatga cattgagtgc ccactgggct ggaagtggga agatgaggaa 120  
 tggccacag acctcaaccg ggctgtcgat gagcaagggt ggcagcatgt ggaacctggc 180  
 gagccccatc cccggcaagc tctcaagcca tgcat 215

<210> 59  
 <211> 246  
 <212> DNA  
 <213> Homo sapiens

<400> 59  
 agagatgggt ccaggagaga tgggggggaag tgccaagcaa tgagtgaccg gttccccctc 60  
 ccccaggctg ggagtatagc atcaccatcc ccccgagcgg gaagccgaag cactgggtcc 120  
 ctgctgagaa gatgtactac acacaccgac ggcgccgctg ggtgcgcctg cgcaggaggg 180  
 atctcagcca aatggaagca ctgaaaaagg gtgagccagc aggtggtggg tgggagtgg 240  
 gcctgt 246

<210> 60  
 <211> 253  
 <212> DNA  
 <213> Homo sapiens

<400> 60  
 cttcccaacc gcctctgagt ctgccccttc ttgtgcagca caggcaggcg gaggcggagg 60  
 gcgagggctg ggagtacgcc tctctttttg gctggaagtt ccacctcgag taccgcaaga 120  
 cagatgcctt ccgcccgcgc cgctggcgcc gtgcgatgga gccactggag aagacggggc 180  
 ctgcagctgt gtttggccctt gagggggccc tggatatgtg ggctgcactt gtcctggctt 240  
 gggtagggta tat 253

<210> 61  
 <211> 177

39/68

<212> DNA  
<213> Homo sapiens

<400> 61  
gaatctgcca taaccagctt cgtgtctcca gggcggcgtg atggatgaca agagtgaaga 60  
ttccatgtcc gtctccacct tgagcttcgg tgtgaacaga cccacgattt cctgcatatt 120  
cgactgtaag taggcttcga ggcctctatg ggggtgataag ggtgtgtcac cttatgc 177

<210> 62  
<211> 181  
<212> DNA  
<213> Homo sapiens

<400> 62  
aaccactcca gccactcact ctggcacctc tgttttttcc cttgggtgaag atgggaaccg 60  
ctaccatcta cgctgctaca tgtaccaggc ccgggacctg gctgcgatgg acaaggactc 120  
tttttctggt aggtgggaga gaggcaggag agtcagagac tgtgggctga gatctgggaa 180  
t 181

<210> 63  
<211> 319  
<212> DNA  
<213> Homo sapiens

<400> 63  
ccccacatgg ctctggagaa gacatctctc agggtccttg ctgtgtaatg tctccctcc 60  
ccctctggcc atgcagatcc ctatgccatc gtctccttcc tgcaccagag ccagaagacg 120  
gtggtggtga agaaccacct taaccccacc tgggaccaga cgctcatctt ctacgagatc 180  
gagatctttg gcgagccggc cacagttgct gagcaaccgc ccagcattgt ggtggagctg 240  
tacgaccatg acacttatgt gagtctgccc agctcctgcc tcgtccctc acagggaggg 300  
accatgtgca aaggtgggg 319

<210> 64  
<211> 249  
<212> DNA  
<213> Homo sapiens

<400> 64  
gccctgggta agggatgctg attcttgtct ctctacgctt ggtctagggt gcagacgagt 60  
ttatgggtcg ctgcatctgt caaccgagtc tggaaacggat gccacggctg gcctgggtcc 120  
cactgacgag gggcagccag ccgtcggggg agctgctggc ctcttttgag ctcatccaga 180  
gagagaaggt gaggtggtc tatatccaga tccaggaggc ccaggcagga gtgggggtggg 240  
ggccaacc 249

<210> 65  
<211> 158  
<212> DNA  
<213> Homo sapiens

<400> 65  
cactgacata gtccatgagt gtcattgagg tgatgggggc cttagggtgac aagcacatga 60  
ccagagctct cttttcttca ctccagccgg ccatccacca tattcctggt tttgaggtaa 120  
gtcttgctct gacctttcct tcttcaaact gattgcc 158

<210> 66  
<211> 132  
<212> DNA  
<213> Homo sapiens

<400> 66  
ctttttcccc ttccaacccc tctcaccatc tcttgatgtg cacatcccat ggctgtgggc 60  
caggtgcagg agacatcaag gatcctggat gaggtgagct ggcggggccg aggtagaggg 120  
aagtggaagc ca 132

<210> 67  
<211> 216  
<212> DNA

40/68

&lt;213&gt; Homo sapiens

&lt;400&gt; 67

|             |            |            |            |            |            |     |
|-------------|------------|------------|------------|------------|------------|-----|
| tcttctctcc  | acctttgtct | ccattctacc | tgctgtccac | tgacgtctga | ggacacagac | 60  |
| ctgcccctacc | caccacccca | gagggaggcc | aacatctaca | tggttcctca | gaacatcaag | 120 |
| ccagcgctcc  | agcgtaccgc | catcgaggtg | agccgtccgg | gcctgggcgt | gggggctggg | 180 |
| agcagcctgc  | ccttcccctt | cctggcccca | gccttt     |            |            | 216 |

&lt;210&gt; 68

&lt;211&gt; 263

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 68

|             |            |             |            |            |            |     |
|-------------|------------|-------------|------------|------------|------------|-----|
| ccccggcctt  | ctgagccact | ctcctcattc  | tgtgtgctta | gaatcctggc | atggggcctg | 60  |
| cggaacatga  | agagttacca | gctggccaac  | atctcctccc | ccagcctcgt | ggtagagtgt | 120 |
| ggggggccaga | cggtgcagtc | ctgtgtcatc  | aggaacctcc | ggaagaacct | caactttgac | 180 |
| atctgcaccc  | tcttcatgga | agtgggtgagc | cccacctccc | tactgtcccc | ttccagagtc | 240 |
| ctgggggctag | aagttctaca | tgt         |            |            |            | 263 |

&lt;210&gt; 69

&lt;211&gt; 249

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 69

|            |             |            |            |            |            |     |
|------------|-------------|------------|------------|------------|------------|-----|
| caggccagtg | cgttcttctt  | cctccaccca | gatgctgccc | agggaggagc | tctactgccc | 60  |
| ccccatcacc | gtcaagggtca | tcgataaacg | ccagtttggc | cgccggcctg | tggtgggcca | 120 |
| gtgtaccatc | cgctccctgg  | agagcttctt | gtgtgacccc | tactcggcgg | agagtccatc | 180 |
| cccacagggt | ggcccaggta  | ggggaagggg | agatgatggg | caggtcaggg | aagggggagc | 240 |
| ctagggcaa  |             |            |            |            |            | 249 |

&lt;210&gt; 70

&lt;211&gt; 180

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 70

|             |            |             |            |            |             |     |
|-------------|------------|-------------|------------|------------|-------------|-----|
| agggggcgagc | cttttgagag | agccccctgtc | aggcctggat | ggctccctcc | cctgcagacg  | 60  |
| atgtgagcct  | actcagtcct | ggggaagacg  | tgctcatcga | cattgatgac | aaggagcccc  | 120 |
| tcatecccat  | ccaggtagga | tgggcatcct  | ccaggagggc | ctgggtcacc | tttccccctcc | 180 |

&lt;210&gt; 71

&lt;211&gt; 211

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 71

|            |             |            |            |            |            |     |
|------------|-------------|------------|------------|------------|------------|-----|
| tgctgcttgg | cgagtcctgt  | ttctgaaatg | gtctctttct | ttctaccac  | tcaggaggaa | 60  |
| gagttcatcg | attgggtggag | caaattcttt | gcctccatag | gggagaggga | aaagtgcggc | 120 |
| tcctacctgg | agaaggattt  | tgacaccctg | aaggtaaggc | ctctcttcag | tctgacagtc | 180 |
| ggtgtgtgtg | tgcgtgctgg  | gcagtgggag | a          |            |            | 211 |

&lt;210&gt; 72

&lt;211&gt; 235

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 72

|            |            |            |            |            |            |     |
|------------|------------|------------|------------|------------|------------|-----|
| gttctacttt | ctttctgtct | cttgtcccct | cctctaattc | ccatgtgtgg | caggtctatg | 60  |
| acacacagct | ggagaatgtg | gaggcctttg | agggcctgtc | tgacttttgt | aacaccttca | 120 |
| agctgtaccg | gggcaagacg | caggaggaga | cagaagatcc | atctgtgatt | ggtgaattta | 180 |
| aggtaaatcc | tcgaagacgt | ccctaaccca | ggtgggccta | agactgtggg | gttgg      | 235 |

&lt;210&gt; 73

&lt;211&gt; 268

&lt;212&gt; DNA

41/68

&lt;213&gt; Homo sapiens

&lt;400&gt; 73

|            |            |            |             |             |             |     |
|------------|------------|------------|-------------|-------------|-------------|-----|
| ggggacacag | ccaaaccata | tcaacaatga | tgataaaaata | aaattaaccc  | ttcctttcttt | 60  |
| tcagggcctc | ttcaaaattt | atccccctcc | agaagaccca  | gccatcccca  | tgcccccag   | 120 |
| acagttccac | cagctggccg | cccagggacc | ccaggagtgc  | ttgggtccgta | tctacattgt  | 180 |
| ccgagcattt | ggcctgcagc | ccaaggaccc | caatggaaag  | gtaactttct  | agagccctca  | 240 |
| cctccccaga | gtagcaggct | caggtaca   |             |             |             | 268 |

&lt;210&gt; 74

&lt;211&gt; 200

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 74

|             |            |            |            |            |            |     |
|-------------|------------|------------|------------|------------|------------|-----|
| tttggaaaagt | gttttcacag | aagtgttttg | tctcctcctc | cagtgtgatc | cttacatcaa | 60  |
| gatctccata  | gggaagaaat | cagtgtgtga | ccaggataac | tacatcccct | gcacgttgga | 120 |
| gcccgtattt  | ggaaagtaaa | ttggggcatc | ttgggtcttg | gggtggagga | gccagacagg | 180 |
| ataaccacaca | gtctagtggg |            |            |            |            | 200 |

&lt;210&gt; 75

&lt;211&gt; 263

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 75

|            |            |            |            |            |            |     |
|------------|------------|------------|------------|------------|------------|-----|
| cctgttccct | tgggtgccct | gtgttggtcg | acattcggga | atctgcccct | tctgcagga  | 60  |
| tgttcgagct | gacctgcact | ctgcctcttg | agaaggacct | aaagatcact | ctctatgact | 120 |
| atgacctcct | ctccaaggac | gaaaagatcg | gtgagacggg | cgtcgacctg | gagaacaggc | 180 |
| tgctgtccaa | gtttggggct | cgctgtggac | tcccacagac | ctactgtgtg | tacgtggatg | 240 |
| ggggctggct | gcctgcttct | ctg        |            |            |            | 263 |

&lt;210&gt; 76

&lt;211&gt; 237

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 76

|            |            |            |            |            |            |     |
|------------|------------|------------|------------|------------|------------|-----|
| aagcatctcg | tctatgtctt | gtgcttgctc | ctcagctctg | gaccgaacca | gtggcgggag | 60  |
| cagctccgcc | cctcccagct | cctccacctc | ttctgccagc | agcatagagt | caaggcacct | 120 |
| gtgtaccgga | cagaccgtgt | aatgtttcag | gataaagaat | attccattga | agagataggt | 180 |
| gagctgccac | atgaccccaa | accatggtgg | gctctcgctg | tatccctccc | tctctca    | 237 |

&lt;210&gt; 77

&lt;211&gt; 245

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 77

|            |            |            |             |            |            |     |
|------------|------------|------------|-------------|------------|------------|-----|
| tctctcgctt | ccccagctcc | tgcaactttt | ttgtgttctc  | tctggggcag | aggctggcag | 60  |
| gatcccaaac | ccacacctgg | gcccagtgga | ggagcgtctg  | gctctgcatg | tgcttcagca | 120 |
| gcagggcctg | gtcccggagc | acgtggagtc | acggccccctc | tacagcccc  | tgacgccaga | 180 |
| catcgagcag | gtaggacctt | acccttggtc | ccagagtcct  | cgaactccag | aagcccaacc | 240 |
| ccagg      |            |            |             |            |            | 245 |

&lt;210&gt; 78

&lt;211&gt; 214

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 78

|            |            |            |            |            |            |     |
|------------|------------|------------|------------|------------|------------|-----|
| ggtgcttggg | aacagctggg | taaatgagaa | gggtggggag | agaacggacc | tgtctccgca | 60  |
| ggggaagctg | gggaagctgc | agatgtgggt | cgacctattt | ccgaaggccc | tggggcggcc | 120 |
| tggacctccc | ttcaacatca | ccccacggag | agccagaagg | tgacttccca | gccacaggct | 180 |
| ctgagctggg | ctgaggggtg | gggcgttgca | gcct       |            |            | 214 |



42/68

<210> 79  
 <211> 229  
 <212> DNA  
 <213> Homo sapiens

<400> 79  
 ttcttaaggc cttcccatcc tttggttagga aatctagggtg gattagagtg atacctttcc 60  
 ccagggttttt cctgcgttgt attatctgga ataccagaga tgtgatcctg gatgacctga 120  
 gcctcacggg ggagaagatg agcgacattt atgtgaaagg gtagggagcc agcgtcctct 180  
 tgccgtgtcca gcttcccgcga gctcccgtgc tccctctggg ttgtgcaca 229

<210> 80  
 <211> 261  
 <212> DNA  
 <213> Homo sapiens

<400> 80  
 acgatgtata tactgtgttg gaaatcttaa tgagaactat tctctaaaaa catgtatgtc 60  
 tagttggatg attggccttg aagaacacaa gcaaaagaca gacgtgcatt atcgttccct 120  
 gggagggtgaa ggcaacttca actggagggtt cattttccccc ttcgactacc tgccagctga 180  
 gcaagtctgt accattgcca agaaggctcag tgtccttccg attccctgtg gtgccagcac 240  
 cagggtcttct aaaggttagcc t 261

<210> 81  
 <211> 234  
 <212> DNA  
 <213> Homo sapiens

<400> 81  
 tgccctctctc taacttttgc tcccttgcac cttctctgtt cctcttccgg gtcaggatgc 60  
 cttctggagg ctggacaaga ctgagagcaa aatcccagca cgagtgggtg tccagatctg 120  
 ggacaatgac aagttctcct ttgatgattt tctgggtgatt ttctgggtaa gcgctattgc 180  
 tagaatccca ttctgcacat gggggctgcc ccagaaccca cactgtgtgt ttat 234

<210> 82  
 <211> 297  
 <212> DNA  
 <213> Homo sapiens

<400> 82  
 ggctacaggc tggcagtgat cgagaaaccc ggccaaaaac cacctctctg ttgcaggctc 60  
 cctgcagctc gatctcaacc gcatgcccac gccagccaag acagccaaga agtgctcctt 120  
 ggaccagctg gatgatgctt tccacccaga atgggtttgtg tccctttttg agcagaaaac 180  
 agtgaagggc tgggtggcct gtgtagcaga agagggtgag aagaaaatac tggcggttaag 240  
 tctacttcct ccagccccag tggaggggcat ggggggaagct tcttccatag aaattgt 297

<210> 83  
 <211> 237  
 <212> DNA  
 <213> Homo sapiens

<400> 83  
 cctgggttact ctccaggcca ctgagcagag ccttcgtgccc cctaaccaag tgctctctgt 60  
 cccctcaggg caagctggaa atgaccttg agattgtagc agagagtga catgaggagc 120  
 ggctgtgctg ccagggccgg gatgagccca acatgaaccc taagcttgag gacccaaggt 180  
 cagtgtcccag cccctgagcc ccaatgcccc caggtctggg ggtataggca cagtcca 237

<210> 84  
 <211> 252  
 <212> DNA  
 <213> Homo sapiens

<400> 84  
 ccctagtaaa ggatgccag ttgactccgg gatctcgctt ccaggcgccc cgacacctcc 60  
 ttccctgtgg ttacctcccc atacaagacc atgaagttca tccctgtggc gcgtttccgg 120  
 tgggccatca tccctctcat catcctcttc atcctgctgc tgttccctggc catcttccatc 180  
 tacgccttcc cgggtgagcag gcctgacgac actgtggttg gggaaactctg ggtctaattg 240

43/68

gggagttcat ca

252

<210> 85  
 <211> 391  
 <212> DNA  
 <213> Homo sapiens

<400> 85  
 tggctgtgcc tgccccagtg ggatcaccat ggggtccctgt ctccctccctc cctccagaac 60  
 tatgctgcca tgaagctggg gaagcccttc agctgaggac tctcctgccc tgtagaaggg 120  
 gccgtggggg cccctccagc atgggactgg cctgcctcct ccgcccagct cggcgagctc 180  
 ctccagacct cctaggcctg attgtcctgc cagggtgggc agacagacag atggaccggc 240  
 ccacactccc agagttgcta acatggagct ctgagatcac cccacttcca tcatttcctt 300  
 ctcccccaac ccaacgcttt tttggatcag ctgagacata tttcagtata aaacagttgg 360  
 aaccacaaaa aaaaaaaaaa aaaaaaaaaa a 391

<210> 86  
 <211> 51  
 <212> PRT  
 <213> Homo sapiens

<400> 86  
 Lys Lys Arg Thr Lys Val Ile Lys Asn Ser Val Asn Pro Val Trp Asn  
 1 5 10 15  
 Glu Gly Phe Glu Trp Asp Leu Lys Gly Ile Pro Leu Asp Gln Gly Ser  
 20 25 30  
 Glu Leu His Val Val Val Lys Asp His Glu Thr Met Gly Arg Asn Arg  
 35 40 45  
 Phe Leu Gly  
 50

<210> 87  
 <211> 45  
 <212> PRT  
 <213> Homo sapiens

<400> 87  
 Ser Lys Ile Leu Glu Lys Thr Ala Asn Pro Gln Trp Asn Gln Asn Ile  
 1 5 10 15  
 Thr Leu Pro Ala Met Phe Pro Ser Met Cys Glu Lys Met Arg Ile Arg  
 20 25 30  
 Ile Ile Asp Trp Asp Arg Leu Thr His Asn Asp Ile Val  
 35 40 45

<210> 88  
 <211> 82  
 <212> PRT  
 <213> Homo sapiens

<400> 88  
 Gln Ala Arg Asp Leu Ala Ala Met Asp Lys Asp Ser Phe Ser Asp Pro  
 1 5 10 15  
 Tyr Ala Ile Val Ser Phe Leu His Gln Ser Gln Lys Thr Val Val Val  
 20 25 30  
 Lys Asn Thr Leu Asn Pro Thr Trp Asp Gln Thr Leu Ile Phe Tyr Glu  
 35 40 45  
 Ile Glu Ile Phe Gly Glu Pro Ala Thr Val Ala Glu Gln Pro Pro Ser  
 50 55 60  
 Ile Val Val Glu Leu Tyr Asp His Asp Thr Tyr Gly Ala Asp Glu Phe  
 65 70 75 80  
 Met Gly

<210> 89  
 <211> 79  
 <212> PRT  
 <213> Homo sapiens

44/68

<400> 89  
 Ile Tyr Ile Val Arg Ala Phe Gly Leu Gln Pro Lys Asp Pro Asn Gly  
           1                  5                  10                  15  
 Lys Cys Asp Pro Tyr Ile Lys Ile Ser Ile Gly Lys Lys Ser Val Ser  
                   20                  25                  30  
 Asp Gln Asp Asn Tyr Ile Pro Cys Thr Leu Glu Pro Val Phe Gly Lys  
                   35                  40                  45  
 Met Phe Glu Leu Thr Cys Thr Leu Pro Leu Glu Lys Asp Leu Lys Ile  
           50                  55                  60  
 Thr Leu Tyr Asp Tyr Asp Leu Leu Ser Lys Asp Glu Lys Ile Gly  
           65                  70                  75

<210> 90  
 <211> 152  
 <212> DNA  
 <213> Homo sapiens

<400> 90  
 acgatgtata tactgtgttg gaaatcttaa tgagaactat tctctaaaaa catgtatgtc 60  
 tagttggatg attggctttg aagaacacaa gcaaaagaca gacgtgcatt atcgttccct 120  
 gggaggtgaa ggcaacttca actggaggtt ca 152

<210> 91  
 <211> 56  
 <212> DNA  
 <213> Homo sapiens

<400> 91  
 gtcagtgatcc ttccgattcc ctgtggtgcc agcaccaggg cttctaaagt tagcct 56

<210> 92  
 <211> 55  
 <212> DNA  
 <213> Homo sapiens

<400> 92  
 tgcctctctc taacttttgc tctttgcac cttctctgtt cctcttccgg gtcag 55

<210> 93  
 <211> 68  
 <212> DNA  
 <213> Homo sapiens

<400> 93  
 gtaagcgcta ttgctagaat cccattctgc acatgggggc tgccccagaa cccacactgt 60  
 gtgtttat 68

<210> 94  
 <211> 56  
 <212> DNA  
 <213> Homo sapiens

<400> 94  
 ggctacaggc tggcagtgat cgagaaaccc ggccaaaaac cacctctctg ttgcag 56

<210> 95  
 <211> 62  
 <212> DNA  
 <213> Homo sapiens

<400> 95  
 gtaagtctac ttcctccagc cccagtggag ggcattgggg aagcttcttc catagaaatt 60  
 gt 62

<210> 96  
 <211> 68  
 <212> DNA

45/68

<213> Homo sapiens  
 <400> 96  
 cctggttact ctccaggcca ctgagcagag ccttcgtgcc cctaaccaag tgctctctgt 60  
 cccctcag 68  
 <210> 97  
 <211> 59  
 <212> DNA  
 <213> Homo sapiens  
 <400> 97  
 gtcagtgtccc agcccctgag ccccaatgcc cacaggtctg ggggtatagg cacagtcca 59  
 <210> 98  
 <211> 44  
 <212> DNA  
 <213> Homo sapiens  
 <400> 98  
 ccctagtaaa ggatgtcccag ttgactccgg gatctcgttt ccag 44  
 <210> 99  
 <211> 60  
 <212> DNA  
 <213> Homo sapiens  
 <400> 99  
 gtgagcaggc ctgacgacac tgtggtgggg gaactctggg tctaattggg gagttcatca 60  
 <210> 100  
 <211> 57  
 <212> DNA  
 <213> Homo sapiens  
 <400> 100  
 tggctgtgcc tgccccagtg ggatcaccat gggtcctgt ctctccctc cctccag 57  
 <210> 101  
 <211> 23  
 <212> DNA  
 <213> Homo sapiens  
 <400> 101  
 tctcttctcc tagaggcca tag 23  
 <210> 102  
 <211> 24  
 <212> DNA  
 <213> Homo sapiens  
 <400> 102  
 ctgttctctc ccatcgtctc atgg 24  
 <210> 103  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens  
 <400> 103  
 gctcctcccg tgaccctctg 20  
 <210> 104  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens

|  |    |
|--|----|
| <400> 104<br>gggtcccagc caggagcact g                     | 21 |
| <210> 105<br><211> 24<br><212> DNA<br><213> Homo sapiens |    |
| <400> 105<br>cccctctcac catctcctga tgtg                  | 24 |
| <210> 106<br><211> 25<br><212> DNA<br><213> Homo sapiens |    |
| <400> 106<br>tggtttcacc ttccctctac ctcg                  | 25 |
| <210> 107<br><211> 24<br><212> DNA<br><213> Homo sapiens |    |
| <400> 107<br>tcctttggta ggaaatctag gtgg                  | 24 |
| <210> 108<br><211> 21<br><212> DNA<br><213> Homo sapiens |    |
| <400> 108<br>ggaagctgga caggcaagag g                     | 21 |
| <210> 109<br><211> 27<br><212> DNA<br><213> Homo sapiens |    |
| <400> 109<br>atatactgtg ttggaaatct taatgag               | 27 |
| <210> 110<br><211> 21<br><212> DNA<br><213> Homo sapiens |    |
| <400> 110<br>gctggcacca cagggaatcg g                     | 21 |
| <210> 111<br><211> 25<br><212> DNA<br><213> Homo sapiens |    |
| <400> 111<br>ctttgcttcc ttgcatcctt ctctg                 | 25 |
| <210> 112<br><211> 21<br><212> DNA<br><213> Homo sapiens |    |
| <400> 112<br>agcccccattg tgcagaatgg g                    | 21 |

47/68

<210> 113  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens

<400> 113  
 ggcaagtgcac gagaaacccg g 21

<210> 114  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens

<400> 114  
 catgccctcc actggggctg g 21

<210> 115  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens

<400> 115  
 ggatgcccag ttgactccgg g 21

<210> 116  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens

<400> 116  
 cccaccaca gtgtcgtcag g 21

<210> 117  
 <211> 6240  
 <212> DNA  
 <213> Homo sapiens

<400> 117  
 atgctgaggg tcttcatcct ctatgccgag aacgtccaca cacccgacac cgacatcagc 60  
 gatgcctact gctccgcggt gtttgcaggg gtgaagaaga gaaccaaagt catcaagaac 120  
 agcgtgaacc ctgtatggaa tgagggattt gaatgggacc tcaagggcat cccctggac 180  
 cagggctctg agcttcatgt ggtggcctaaa gaccatgaga cgatggggag gaacagggtc 240  
 ctgggggaag ccaagggtccc actccgagag gtccctcgcca cccctagtct gtccgcccagc 300  
 ttcaatgccc cccgtgctgga caccaagaag cagcccacag gggcctcgct ggtcctgcag 360  
 gtgtcctaca caccgctgcc tggagctgtg cccctgttcc cgccccctac tcctctggag 420  
 ccctccccga ctctgcctga cctggatgta gtggcagaca caggaggaga ggaagacaca 480  
 gaggaccagg gactcactgg agatgaggcg gagccattcc tggatcaaag cggaggcccg 540  
 ggggctccca ccaccccaag gaaactacct tcacgtcctc cgccccacta ccccgggatc 600  
 aaaagaaagc gaagtgcgcc tacatctaga aagctgctgt cagacaaacc gcaggatttc 660  
 cagatcaggg tccaggtgat cgaggggcgc cagctgccgg ggggtgaacat caagcctgtg 720  
 gtcaagggtta ccgctgcagg gcagaccaag cggacgcgga tccacaaggg aaacagccca 780  
 ctcttcaatg agactctttt cttcaacttg tttgactctc ctggggagct gtttgatgag 840  
 cccatcttta tcacggtggt agactctcgt tctctcagga cagatgctct cctcggggag 900  
 ttccggatgg acgtgggcac catttacaga gagccccggc acgcctatct caggaagtgg 960  
 ctgctgctct cagaccctga tgacttctct gctggggcca gaggctacct gaaaacaagc 1020  
 ctttgtgtgc tggggccttg ggacgaagcg cctctggaga gaaaagacct ctctgaagac 1080  
 aaggaggaca ttgaaagcaa cctgctccgg cccacaggcg tagccctgcg aggagcccac 1140  
 ttctgcctga aggtcttccg ggccgaggac ttgccgcaga tggacgatgc cgtgatggac 1200  
 aacgtgaaac agatctttgg cttcgagagt aacaagaaga acttgggtgga cccctttgtg 1260  
 gaggtcagct ttgcggggaa aatgctgtgc agcaagatct tggagaagac ggccaaccct 1320  
 cagtggaacc agaactcac actgcctgcc atgtttccct ccatgtgcga aaaaatgagg 1380  
 attcgtatca tagactggga ccgcctgact cacaatgaca tcgtggctac cacctacctg 1440  
 agtatgtcga aaatctctgc cctggaggga gaaatagaag aggagcctgc aggtgctgtc 1500  
 aagccttcga aagcctcaga cttggatgac tacctgggct tcctccccac ttttgggccc 1560  
 tgctacatca acctctatgg cagtcccaga gagttcacag gcttcccaga cccctacaca 1620  
 gagctcaaca caggcaaggg ggaaggtgtg gcttatcgtg gccggcttct gctctccctg 1680  
 gagaccaagc tgggtggagca cagtgaacag aaggtggagg accttcctgc ggatgacatc 1740

|             |             |             |             |             |            |      |
|-------------|-------------|-------------|-------------|-------------|------------|------|
| ctccgggtg   | agaagtac    | taggaggg    | aagtactccc  | tgtttg      | cttctactca | 1800 |
| gccaccatg   | tcagga      | ggatgatg    | atccagttt   | aggtcagcat  | cgggaactac | 1860 |
| gggaacaagt  | tcgacatg    | ctgcctg     | ctggcctcca  | ccactcagta  | cagccgtgca | 1920 |
| gtctttgacg  | ggtgccacta  | ctactaccta  | ccctggggta  | acgtgaaacc  | tgtggtggtg | 1980 |
| ctgtcatcct  | actgggagga  | catcagccat  | agaatcgaga  | ctcagaacca  | gctgcttggg | 2040 |
| attgctgacc  | ggctggaagc  | tggcctggag  | caggtccacc  | tggccctgaa  | ggcgagtg   | 2100 |
| tccacggagg  | acgtggactc  | gctggtggt   | cagctgacgg  | atgagctcat  | cgcaggctgc | 2160 |
| agccagcctc  | tgggtgacat  | ccatgagaca  | ccctctgcca  | cccacctgga  | ccagtacctg | 2220 |
| taccagctgc  | gcacccatca  | cctgagccaa  | atcactgagg  | ctgccctggc  | cctgaagctc | 2280 |
| ggccacagtg  | agctccctgc  | agctctggag  | caggcggagg  | actggctcct  | gcgtctgcgt | 2340 |
| gccctggcag  | aggagcccca  | gaacagcctg  | ccggacatcg  | tcactctggat | gctgcaggga | 2400 |
| gacaagcgtg  | tgccatacca  | gcggtgccc   | gcccaccaag  | tcctcttctc  | ccggcggggt | 2460 |
| gccaactact  | gtggcaagaa  | ttgtgggaag  | ctacagacaa  | tctttctgaa  | atatccgatg | 2520 |
| gagaaggtgc  | ctggcgcccc  | gatgccagtg  | cagatacggg  | tcaagctgtg  | gtttgggctc | 2580 |
| tctgtggatg  | agaaggagtt  | caaccagttt  | gctgagggga  | agctgtctgt  | ctttgctgaa | 2640 |
| acctatgaga  | accgagactaa | gttggccctt  | gttgggaact  | ggggcacaac  | gggcctcacc | 2700 |
| tacccaagt   | ttcttgacgt  | cacgggcaag  | atcaagctac  | ccaaggacag  | cttcgcctcc | 2760 |
| tcggccggct  | ggacctgggc  | tggagattgg  | ttcgtgtgtc  | cggagaagac  | tctgctccat | 2820 |
| gacatggacg  | ccggtcacct  | gagcttcgtg  | gaagaggtgt  | ttgagaacca  | gacccggctt | 2880 |
| cccgaggggc  | agtggatcta  | catgagtgc   | aactaacccg  | atgtgaacgg  | ggagaaggtg | 2940 |
| cttcccaagg  | atgacattga  | gtgccactg   | ggctgggaagt | gggaagatga  | ggaatggtcc | 3000 |
| acagacctca  | accgggtgt   | cgatgagcaa  | ggctgggagt  | atagcatcac  | catccccccg | 3060 |
| gagcggaaagc | cgaagcactg  | ggtccctgct  | gagaagatgt  | actacacaca  | ccgacggcgg | 3120 |
| cgctgggtgc  | gcctgcgcag  | gagggatctc  | agccaaatgg  | aagcactgaa  | aaggcacagg | 3180 |
| caggcggagg  | cggagggcga  | gggctgggag  | tacgcctctc  | tttttggtcg  | gaagttccac | 3240 |
| ctcgagtacc  | gcaagacaga  | tgccctccgc  | cgccgcgctc  | ggcgccgtcg  | catggagcca | 3300 |
| ctggagaaga  | cggggcctgc  | agctgtgttt  | gcccttgagg  | gggccctggg  | cggcgtgatg | 3360 |
| gatgacaaga  | gtgaagattc  | catgtccgtc  | tccaccttga  | gcttcgggtg  | gaacagaccc | 3420 |
| acgatttcct  | gcatattcga  | ctatgggaac  | cgctaccatc  | tacgtctgta  | catgtaccag | 3480 |
| gcccgggacc  | tggtctgcgt  | ggacaaggac  | tctttttctg  | atccctatgc  | catcgtctcc | 3540 |
| ttcctgcacc  | agagccagaa  | gacggtggtg  | gtgaagaaca  | cccttaaccc  | cacctgggac | 3600 |
| cagacgtca   | tctttctacga | gatcgagatc  | tttgccgagc  | cggccacagt  | tgtgagcaa  | 3660 |
| ccgcccagca  | ttgtggtgga  | gctgtacgac  | catgacactt  | atggtgcaga  | cgagtttatg | 3720 |
| ggtcgctgca  | tctgtcaacc  | gagctctggaa | cggatgccac  | ggctggcctg  | gttcccactg | 3780 |
| acgaggggca  | gccagccgtc  | gggggagctg  | ctggcctctt  | ttgagctcat  | ccagagagag | 3840 |
| aagccggcca  | tcacacatat  | tccctggttt  | gaggtgcagg  | agacatcaag  | gatcctggat | 3900 |
| gagctcgagg  | acacagacct  | gccctaccca  | ccaccccgaga | gggaggccaa  | catctacatg | 3960 |
| gttccctcaga | acatcaagcc  | agcgtccag   | cgtaccgcga  | tcgagatcct  | ggcatggggc | 4020 |
| ctgcggaaca  | tgaagagtta  | ccagctggcc  | aacatctcct  | ccccagcct   | cgtggtagag | 4080 |
| tgtggggggc  | agacggtgca  | gtcctgtgtc  | atcaggaacc  | tccggaagaa  | ccccacttt  | 4140 |
| gacatctgca  | ccctcttcat  | ggaagtgatg  | ctgcccaggg  | aggagctcta  | ctgccccccc | 4200 |
| atcacctgca  | aggtcatcga  | taaccgccag  | tttgcccgcc  | ggcctgtggt  | gggccagtg  | 4260 |
| accatccgct  | ccctggagag  | cttcctgtgt  | gacccctact  | cggcggagag  | tccatcccc  | 4320 |
| cagggtggcc  | cagacgatgt  | gagcctactc  | agtccctggg  | aagacgtgct  | catcgacatt | 4380 |
| gatgacaagg  | agcccccat   | cccatccag   | gaggaagagt  | tcactcgattg | gtggagcaaa | 4440 |
| ttctttgcct  | ccatagggga  | gagggaaaag  | tgccgctcct  | acctggagaa  | ggattttgac | 4500 |
| accctgaagg  | tctatgacac  | acagctggag  | aatgtggagg  | cctttgaggg  | cctgtctgac | 4560 |
| ttttgtaaca  | ccttcaagct  | gtaccggggc  | aagacgcagg  | aggagacaga  | agatccatct | 4620 |
| gtgattggtg  | aatttaaggg  | cctcttcaaa  | atcttatccc  | tcccagaaga  | cccagccatc | 4680 |
| cccattgccc  | caagacagtt  | ccaccagctg  | gcccggcagg  | gacccagga   | gtgcttggtc | 4740 |
| cgtatctaca  | ttgtccgagc  | atcttgccctg | cagcccaagg  | accccaatgg  | aaagtgtgat | 4800 |
| ccttacatca  | agatctccat  | agggaaagaa  | tcagtgagtg  | accaggataa  | ctacatcccc | 4860 |
| tgcacgctgg  | agcccgtatt  | tggaaagatg  | ttcgagctga  | cctgcactct  | gcctctggag | 4920 |
| aaggacctaa  | agatcactct  | ctatgactat  | gacctcctct  | ccaaggacga  | aaagatcggt | 4980 |
| gagacggctg  | tcgacctgga  | gaacaggctg  | ctgtccaagt  | ttggggctcg  | ctgtggactc | 5040 |
| ccacagacct  | actgtgtctc  | tggaccgaac  | cagtgccggg  | accagctccg  | cccctcccag | 5100 |
| ctcctccacc  | tcttctgcca  | gcagcataga  | gtcaaggcac  | ctgtgtaccg  | gacagaccgt | 5160 |
| gtaatgtttc  | aggataaaga  | atattccatt  | gaagagatag  | aggctggcag  | gatcccaaac | 5220 |
| ccacacctgg  | gcccagtgga  | ggagcgtctg  | tgctctgcatg | tgcttcagca  | gcaggccctg | 5280 |
| gtcccggagc  | acgtggagtc  | acggccctc   | tacagccccc  | tgcagccaga  | catcgagcag | 5340 |
| gggaagctgc  | agatgtgggt  | cgacctat    | ccgaaggccc  | tggggcgggc  | tggacctccc | 5400 |
| ttcaacatca  | ccccacggag  | agccagaagg  | tttttcctgc  | gttgatttat  | ctggaatacc | 5460 |
| agagatgtga  | tcttgatga   | cctgagcctc  | acgggggaga  | agatgagcga  | catttatgtg | 5520 |
| aaagggttga  | tgattggctt  | tgaagaacac  | aagcaaaaga  | cagacgtgca  | ttatcgttcc | 5580 |
| ctgggaggtg  | aggcaacctt  | caactggagg  | ttcattttcc  | ccttcgacta  | cctgccagct | 5640 |
| gagcaagctc  | gtaccattgc  | caagaaggat  | gccttctgga  | ggctggacaa  | gactgagagc | 5700 |
| aaaatcccag  | cacgagtgg   | gttccagatc  | tgggacaatg  | acaagttctc  | ctttgatgat | 5760 |

49/68

|            |            |             |            |            |            |      |
|------------|------------|-------------|------------|------------|------------|------|
| tttctgggct | ccctgcagct | cgatctcaac  | cgcatgcccc | agccagccaa | gacagccaag | 5820 |
| aagtgtcct  | tgaccagct  | ggatgatgct  | ttccaccag  | aatggtttgt | gtcccttttt | 5880 |
| gagcagaaaa | cagtgaagg  | ctgggtggccc | tgtgtagcag | aagagggtga | gaagaaaata | 5940 |
| ctggcgggca | agctggaaat | gaccttggag  | attgtagcag | agagttagca | tgaggagcgg | 6000 |
| cctgctggcc | agggccggga | tgagcccaac  | atgaacccta | agcttgagga | cccaaggcgc | 6060 |
| cccgaacac  | ccttcctgtg | gtttacctcc  | ccatacaaga | ccatgaagtt | catcctgtgg | 6120 |
| cggcgtttcc | ggtgggcat  | catcctcttc  | atcatcctct | tcacctctgt | gctgttcctg | 6180 |
| gccatcttca | tctacgcctt | cccgaactat  | gctgccatga | agctggtgaa | gcccttcagc | 6240 |

<210> 118  
 <211> 13  
 <212> DNA  
 <213> Homo sapiens

<400> 118  
 cgcaagcatg ctg 13

<210> 119  
 <211> 12  
 <212> DNA  
 <213> Homo sapiens

<400> 119  
 gagacgatgg gg 12

<210> 120  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens

<400> 120  
 gatctaacc tgetgctcac c 21

<210> 121  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens

<400> 121  
 ctggtgtgtt gcagagcgct g 21

<210> 122  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens

<400> 122  
 cctctcttct gctgtcttca g 21

<210> 123  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens

<400> 123  
 tgtgtctggt tcaccttcgt g 21

<210> 124  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens

<400> 124  
 tccaaataga aatgcctgaa c 21

<210> 125  
 <211> 21



50/68

<212> DNA  
<213> Homo sapiens  
  
<400> 125 21  
aggtatcacc tccaagtgtt g  
  
<210> 126  
<211> 21  
<212> DNA  
<213> Homo sapiens  
  
<400> 126 21  
taccagcttc agagctccct g  
  
<210> 127  
<211> 19  
<212> DNA  
<213> Homo sapiens  
  
<400> 127 19  
ttgatcaggg tgctcttgg  
  
<210> 128  
<211> 20  
<212> DNA  
<213> Homo sapiens  
  
<400> 128 20  
ggagaattgc ttgaaccag  
  
<210> 129  
<211> 22  
<212> DNA  
<213> Homo sapiens  
  
<400> 129 22  
tggctaata tggtaaacat tt  
  
<210> 130  
<211> 21  
<212> DNA  
<213> Homo sapiens  
  
<400> 130 21  
gaccacaag cggcgctcg g  
  
<210> 131  
<211> 21  
<212> DNA  
<213> Homo sapiens  
  
<400> 131 21  
gaccccggcg aggggtggtcg g  
  
<210> 132  
<211> 24  
<212> DNA  
<213> Homo sapiens  
  
<400> 132 24  
tgtctctcca ttctcccttt tgtg  
  
<210> 133  
<211> 24  
<212> DNA  
<213> Homo sapiens

|  |    |
|--|----|
| <400> 133<br>aggacactgc tgagaaggca cctc                  | 24 |
| <210> 134<br><211> 21<br><212> DNA<br><213> Homo sapiens |    |
| <400> 134<br>agtgccctgg tggcacgaag g                     | 21 |
| <210> 135<br><211> 24<br><212> DNA<br><213> Homo sapiens |    |
| <400> 135<br>cctacctgca ccttcaagcc atgg                  | 24 |
| <210> 136<br><211> 23<br><212> DNA<br><213> Homo sapiens |    |
| <400> 136<br>cagaagagcc aggggtgcctt agg                  | 23 |
| <210> 137<br><211> 24<br><212> DNA<br><213> Homo sapiens |    |
| <400> 137<br>ccttggacct taacctggca gagg                  | 24 |
| <210> 138<br><211> 21<br><212> DNA<br><213> Homo sapiens |    |
| <400> 138<br>cgaggccagc gcaccaacct g                     | 21 |
| <210> 139<br><211> 22<br><212> DNA<br><213> Homo sapiens |    |
| <400> 139<br>actgccggcc attcttgctg gg                    | 22 |
| <210> 140<br><211> 21<br><212> DNA<br><213> Homo sapiens |    |
| <400> 140<br>ccaggcctca ttagggccct c                     | 21 |
| <210> 141<br><211> 22<br><212> DNA<br><213> Homo sapiens |    |
| <400> 141<br>ctgaagagga gcctggggtc ag                    | 22 |

52/68

<210> 142  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 142  
ctgagatttc tgactcttgg ggtg 24

<210> 143  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 143  
aaggttctgc cctcatgccc catg 24

<210> 144  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 144  
ctggcctgag ggatcagcag g 21

<210> 145  
<211> 23  
<212> DNA  
<213> Homo sapiens

<400> 145  
gtgcatacat acagcccacg gag 23

<210> 146  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 146  
gagctattgg gttggccgtg tggg 24

<210> 147  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 147  
accaacacgg agaagtgaga actg 24

<210> 148  
<211> 26  
<212> DNA  
<213> Homo sapiens

<400> 148  
ccacacttta tttaacgctt tggcgg 26

<210> 149  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 149  
cagaacaaa atgcaaggat acgg 24

<210> 150  
<211> 25  
<212> DNA

53/68

<213> Homo sapiens  
 <400> 150  
 cttctgattc tgggatcacc aaagg 25  
 <210> 151  
 <211> 22  
 <212> DNA  
 <213> Homo sapiens  
 <400> 151  
 ggaccgtaag gaagacccag gg 22  
 <210> 152  
 <211> 24  
 <212> DNA  
 <213> Homo sapiens  
 <400> 152  
 cctgtgctca ggagcgcacg aagg 24  
 <210> 153  
 <211> 22  
 <212> DNA  
 <213> Homo sapiens  
 <400> 153  
 gcagacctcc caccgaagg cg 22  
 <210> 154  
 <211> 24  
 <212> DNA  
 <213> Homo sapiens  
 <400> 154  
 gagacagatg ggggacagtc aggg 24  
 <210> 155  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens  
 <400> 155  
 cctcccgcaga gaaccctcct g 21  
 <210> 156  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens  
 <400> 156  
 gggagcccag agtcccatg g 21  
 <210> 157  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens  
 <400> 157  
 gggcctcctt gggtttgctg g 21  
 <210> 158  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens

|  |    |
|--|----|
| <400> 158<br>gcctccccag catcctgccg g                     | 21 |
| <210> 159<br><211> 24<br><212> DNA<br><213> Homo sapiens |    |
| <400> 159<br>tcactgagcc gaatgaaact gagg                  | 24 |
| <210> 160<br><211> 24<br><212> DNA<br><213> Homo sapiens |    |
| <400> 160<br>tgtggcctga gttcctttcc tgtg                  | 24 |
| <210> 161<br><211> 24<br><212> DNA<br><213> Homo sapiens |    |
| <400> 161<br>ggtcaaaggg cagaacgaag aggg                  | 24 |
| <210> 162<br><211> 21<br><212> DNA<br><213> Homo sapiens |    |
| <400> 162<br>cccgtccttc tcccagccat g                     | 21 |
| <210> 163<br><211> 21<br><212> DNA<br><213> Homo sapiens |    |
| <400> 163<br>ctcccctggt tgtccccaag g                     | 21 |
| <210> 164<br><211> 24<br><212> DNA<br><213> Homo sapiens |    |
| <400> 164<br>cgacccctct gattgccact tgtg                  | 24 |
| <210> 165<br><211> 21<br><212> DNA<br><213> Homo sapiens |    |
| <400> 165<br>ggcatcctgc ccttgccagg g                     | 21 |
| <210> 166<br><211> 20<br><212> DNA<br><213> Homo sapiens |    |
| <400> 166<br>tctgtctccc ctgctccttg                       | 20 |

55/68

<210> 167  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens  
  
 <400> 167  
 cttccctgcc ccgacgccca g 21  
  
 <210> 168  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens  
  
 <400> 168  
 cagcgctcag gcccgtctct c 21  
  
 <210> 169  
 <211> 24  
 <212> DNA  
 <213> Homo sapiens  
  
 <400> 169  
 tgcataaggca tgtgcagctt tggg 24  
  
 <210> 170  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens  
  
 <400> 170  
 catgcaccct ctgccctgtg g 21  
  
 <210> 171  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens  
  
 <400> 171  
 agttgagcca ggagaggtgg g 21  
  
 <210> 172  
 <211> 24  
 <212> DNA  
 <213> Homo sapiens  
  
 <400> 172  
 catcaggcgc attccatctg tccg 24  
  
 <210> 173  
 <211> 24  
 <212> DNA  
 <213> Homo sapiens  
  
 <400> 173  
 agcaggagag cagaagaaga aagg 24  
  
 <210> 174  
 <211> 22  
 <212> DNA  
 <213> Homo sapiens  
  
 <400> 174  
 gtgtgtcacc atccccaccc cg 22  
  
 <210> 175  
 <211> 25  
 <212> DNA

56/68

<213> Homo sapiens  
 <400> 175  
 caagagatgg gagaaaggcc ttatg 25  
 <210> 176  
 <211> 23  
 <212> DNA  
 <213> Homo sapiens  
 <400> 176  
 ctgggacatc cggatcctga agg 23  
 <210> 177  
 <211> 22  
 <212> DNA  
 <213> Homo sapiens  
 <400> 177  
 tccaggtagt gggaggcaga gg 22  
 <210> 178  
 <211> 24  
 <212> DNA  
 <213> Homo sapiens  
 <400> 178  
 tcccactacc tggagctgcc ttgg 24  
 <210> 179  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens  
 <400> 179  
 ggctctcccc agcctccct g 21  
 <210> 180  
 <211> 24  
 <212> DNA  
 <213> Homo sapiens  
 <400> 180  
 cagagcagca gagactctga ccag 24  
 <210> 181  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens  
 <400> 181  
 tagacccac ctgcccctga g 21  
 <210> 182  
 <211> 24  
 <212> DNA  
 <213> Homo sapiens  
 <400> 182  
 tcctctcatt gcttgccctgt tcgg 24  
 <210> 183  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens

57/68

<400> 183  
ttgagagctt gccggggatg g 21  
    <210> 184  
    <211> 24  
    <212> DNA  
    <213> Homo sapiens

<400> 184  
aagtgccaaag caatgagtga ccgg 24  
    <210> 185  
    <211> 21  
    <212> DNA  
    <213> Homo sapiens

<400> 185  
ctcactccca cccaccacct g 21  
    <210> 186  
    <211> 21  
    <212> DNA  
    <213> Homo sapiens

<400> 186  
cccaccggcc tctgagtctg c 21  
    <210> 187  
    <211> 24  
    <212> DNA  
    <213> Homo sapiens

<400> 187  
accctaccca agccaggaca agtg 24  
    <210> 188  
    <211> 24  
    <212> DNA  
    <213> Homo sapiens

<400> 188  
gaatctgccca taaccagctt cgtg 24  
    <210> 189  
    <211> 24  
    <212> DNA  
    <213> Homo sapiens

<400> 189  
tatcacccca tagaggcctc gaag 24  
    <210> 190  
    <211> 24  
    <212> DNA  
    <213> Homo sapiens

<400> 190  
cagccactca ctctggcacc tctg 24  
    <210> 191  
    <211> 24  
    <212> DNA  
    <213> Homo sapiens

<400> 191  
agcccacagt ctctgactct cctg 24



58/68

<210> 192  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 192  
acatctctca gggccctgc tgtg 24

<210> 193  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 193  
cctgtgaggg gacgaggcag g 21

<210> 194  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 194  
gccctgggta agggatgctg attc 24

<210> 195  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 195  
cctgcctggg cctcctggat c 21

<210> 196  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 196  
gagggtgatg ggggccttag g 21

<210> 197  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 197  
gcaatcagtt tgaagaagga aagg 24

<210> 198  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 198  
cacctttgtc tccattctac ctgc 24

<210> 199  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 199  
ctcccagccc ccacgccag g 21

<210> 200  
<211> 24  
<212> DNA

<213> Homo sapiens  
<400> 200  
ctgagccact ctcctcattc tgtg 24  
<210> 201  
<211> 21  
<212> DNA  
<213> Homo sapiens  
<400> 201  
tggaagggga cagtagggag g 21  
<210> 202  
<211> 22  
<212> DNA  
<213> Homo sapiens  
<400> 202  
ggccagtgcg ttcttcctcc tc 22  
<210> 203  
<211> 22  
<212> DNA  
<213> Homo sapiens  
<400> 203  
tccctgacct gcccatcatc tc 22  
<210> 204  
<211> 21  
<212> DNA  
<213> Homo sapiens  
<400> 204  
gcccctgtca ggccctggatg g 21  
<210> 205  
<211> 21  
<212> DNA  
<213> Homo sapiens  
<400> 205  
tgacccaggc ctccctggag g 21  
<210> 206  
<211> 24  
<212> DNA  
<213> Homo sapiens  
<400> 206  
ctgaaatggt ctctttcttt ctac 24  
<210> 207  
<211> 24  
<212> DNA  
<213> Homo sapiens  
<400> 207  
cacaccgact gtcagactga agag 24  
<210> 208  
<211> 24  
<212> DNA  
<213> Homo sapiens

|  |    |
|--|----|
| <400> 208<br>ttgtcccctc ctctaattccc catg                 | 24 |
| <210> 209<br><211> 21<br><212> DNA<br><213> Homo sapiens |    |
| <400> 209<br>ggggttaggga cgtcttcgag g                    | 21 |
| <210> 210<br><211> 22<br><212> DNA<br><213> Homo sapiens |    |
| <400> 210<br>cagccaaacc atatcaacaa tg                    | 22 |
| <210> 211<br><211> 21<br><212> DNA<br><213> Homo sapiens |    |
| <400> 211<br>ctggggaggt gagggctcta g                     | 21 |
| <210> 212<br><211> 21<br><212> DNA<br><213> Homo sapiens |    |
| <400> 212<br>gaagtgtttt gtctcctcct c                     | 21 |
| <210> 213<br><211> 20<br><212> DNA<br><213> Homo sapiens |    |
| <400> 213<br>gcaggcagcc agcccccatc                       | 20 |
| <210> 214<br><211> 21<br><212> DNA<br><213> Homo sapiens |    |
| <400> 214<br>gggtgccctg tgttggtga c                      | 21 |
| <210> 215<br><211> 20<br><212> DNA<br><213> Homo sapiens |    |
| <400> 215<br>gcaggcagcc agcccccatc                       | 20 |
| <210> 216<br><211> 24<br><212> DNA<br><213> Homo sapiens |    |
| <400> 216<br>ctcgtctatg tcttggtgctt gctc                 | 24 |

61/68

<210> 217  
<211> 23  
<212> DNA  
<213> Homo sapiens  
  
<400> 217  
caccatgggtt tggggtcacg tgg 23  
  
<210> 218  
<211> 21  
<212> DNA  
<213> Homo sapiens  
  
<400> 218  
tctcgcttcc ccagctcctg c 21  
  
<210> 219  
<211> 22  
<212> DNA  
<213> Homo sapiens  
  
<400> 219  
tctggagttc gaggactctg gg 22  
  
<210> 220  
<211> 21  
<212> DNA  
<213> Homo sapiens  
  
<400> 220  
agaagggtgg ggagagaacg g 21  
  
<210> 221  
<211> 21  
<212> DNA  
<213> Homo sapiens  
  
<400> 221  
cagctcagag cctgtggctg g 21  
  
<210> 222  
<211> 24  
<212> DNA  
<213> Homo sapiens  
  
<400> 222  
aaggccttcc catcctttgg tagg 24  
  
<210> 223  
<211> 21  
<212> DNA  
<213> Homo sapiens  
  
<400> 223  
acaacccaga gggagcacgg g 21  
  
<210> 224  
<211> 25  
<212> DNA  
<213> Homo sapiens  
  
<400> 224  
gttgacgatg tatatactgt gttgg 25  
  
<210> 225  
<211> 25  
<212> DNA

62/68

```

<213> Homo sapiens
<400> 225
gcctctctct aactttgctt ccttg                25
<210> 226
<211> 24
<212> DNA
<213> Homo sapiens
<400> 226
ggctacaggc tggcagtgat cgag                24
<210> 227
<211> 21
<212> DNA
<213> Homo sapiens
<400> 227
ttcccccatg ccctccactg g                21
<210> 228
<211> 24
<212> DNA
<213> Homo sapiens
<400> 228
agccttcgtg cccctaacca agtg                24
<210> 229
<211> 21
<212> DNA
<213> Homo sapiens
<400> 229
ctgtgggcat tggggctcag g                21
<210> 230
<211> 20
<212> DNA
<213> Homo sapiens
<400> 230
gccccagtg g gatcaccatg                20
<210> 231
<211> 21
<212> DNA
<213> Homo sapiens
<400> 231
atgctggagg ggaccccacg g                21
<210> 232
<211> 3671
<212> DNA
<213> Homo sapiens
<220>
<221> CDS
<222> (418)...(3381)
<400> 232
tcctggttca agcgattctc tggcctcagc ctcccgagta gctgggatta caggcatgct    60
ccaccaagcc cgggtaattt tgtattttta atagagacgg ggttttgcca tggtggtcag    120
gctgggtctcg aactcctgac ctcaggtgat ctgcccacct tggcctccca acgtgctgag    180
attacaggca tgagtcactg tgcccggcag agatggtcta attcatatga aagaactctg    240

```

63/68

|   |                 |            |            |             |             |     |
|---|-----------------|------------|------------|-------------|-------------|-----|
| aaaaaagtag  | aaagtgattt      | tctaaaataa | ggtacaaata | attaatgtaa  | gcataatcac  | 300 |
| ctaacccttgt   | ggaatttttt      | ttttttgaga | agcaaattgc | aaattttgtga | tagatctaaa  | 360 |
| ggagattgac  | taagagggtg      | accatctgga | aatgacgtca | tgtgagaatg  | gttaaag atg | 420 |
|   |                 |            |            |             | Met         |     |
|   |                 |            |            |             | 1           |     |
| ctc ggg aga ttg agc cta gag aaa gga aga ttt gtg aac cca gga ggc | 468             |            |            |             |             |     |
| Leu Gly Arg Leu Ser Leu Glu Lys Gly Arg Phe Val Asn Pro Gly Gly |                 |            |            |             |             |     |
|   | 5 10 15         |            |            |             |             |     |
| aga ggt aga gat cca gga gag ggc ggc gtg atg gat gac aag agt gaa | 516             |            |            |             |             |     |
| Arg Gly Arg Asp Pro Gly Glu Gly Gly Val Met Asp Asp Lys Ser Glu |                 |            |            |             |             |     |
|   | 20 25 30        |            |            |             |             |     |
| gat tcc atg tcc gtc tcc acc ttg agc ttc ggt gtg aac aga ccc acg | 564             |            |            |             |             |     |
| Asp Ser Met Ser Val Ser Thr Leu Ser Phe Gly Val Asn Arg Pro Thr |                 |            |            |             |             |     |
|   | 35 40 45        |            |            |             |             |     |
| att tcc tgc ata ttc gac tat ggg aac cgc tac cat cta cgc tgc tac | 612             |            |            |             |             |     |
| Ile Ser Cys Ile Phe Asp Tyr Gly Asn Arg Tyr His Leu Arg Cys Tyr |                 |            |            |             |             |     |
|   | 50 55 60 65     |            |            |             |             |     |
| atg tac cag gcc cgg gac ctg gct gcg atg gac aag gac tct ttt tct | 660             |            |            |             |             |     |
| Met Tyr Gln Ala Arg Asp Leu Ala Ala Met Asp Lys Asp Ser Phe Ser |                 |            |            |             |             |     |
|   | 70 75 80        |            |            |             |             |     |
| gat ccc tat gcc atc gtc tcc ttc ctg cac cag agc cag aag acg gtg | 708             |            |            |             |             |     |
| Asp Pro Tyr Ala Ile Val Ser Phe Leu His Gln Ser Gln Lys Thr Val |                 |            |            |             |             |     |
|   | 85 90 95        |            |            |             |             |     |
| gtg gtg aag aac acc ctt aac ccc acc tgg gac cag acg ctc atc ttc | 756             |            |            |             |             |     |
| Val Val Lys Asn Thr Leu Asn Pro Thr Trp Asp Gln Thr Leu Ile Phe |                 |            |            |             |             |     |
|   | 100 105 110     |            |            |             |             |     |
| tac gag atc gag atc ttt ggc gag ccg gcc aca gtt gct gag caa ccg | 804             |            |            |             |             |     |
| Tyr Glu Ile Glu Ile Phe Gly Glu Pro Ala Thr Val Ala Glu Gln Pro |                 |            |            |             |             |     |
|   | 115 120 125     |            |            |             |             |     |
| ccc agc att gtg gtg gag ctg tac gac cat gac act tat ggt gca gac | 852             |            |            |             |             |     |
| Pro Ser Ile Val Val Glu Leu Tyr Asp His Asp Thr Tyr Gly Ala Asp |                 |            |            |             |             |     |
|   | 130 135 140 145 |            |            |             |             |     |
| gag ttt atg ggt cgc tgc atc tgt caa ccg agt ctg gaa cgg atg cca | 900             |            |            |             |             |     |
| Glu Phe Met Gly Arg Cys Ile Cys Gln Pro Ser Leu Glu Arg Met Pro |                 |            |            |             |             |     |
|   | 150 155 160     |            |            |             |             |     |
| cgg ctg gcc tgg ttc cca ctg acg agg ggc agc cag ccg tcg ggg gag | 948             |            |            |             |             |     |
| Arg Leu Ala Trp Phe Pro Leu Thr Arg Gly Ser Gln Pro Ser Gly Glu |                 |            |            |             |             |     |
|   | 165 170 175     |            |            |             |             |     |
| ctg ctg gcc tct ttt gag ctc atc cag aga gag aag ccg gcc atc cac | 996             |            |            |             |             |     |
| Leu Leu Ala Ser Phe Glu Leu Ile Gln Arg Glu Lys Pro Ala Ile His |                 |            |            |             |             |     |
|   | 180 185 190     |            |            |             |             |     |
| cat att cct ggt ttt gag gtg cag gag aca tca agg atc ctg gat gag | 1044            |            |            |             |             |     |
| His Ile Pro Gly Phe Glu Val Gln Glu Thr Ser Arg Ile Leu Asp Glu |                 |            |            |             |             |     |
|   | 195 200 205     |            |            |             |             |     |
| tct gag gac aca gac ctg ccc tac cca cca ccc cag agg gag gcc aac | 1092            |            |            |             |             |     |
| Ser Glu Asp Thr Asp Leu Pro Tyr Pro Pro Pro Gln Arg Glu Ala Asn |                 |            |            |             |             |     |
|   | 210 215 220 225 |            |            |             |             |     |
| atc tac atg gtt cct cag aac atc aag cca gcg ctc cag cgt acc gcc | 1140            |            |            |             |             |     |
| Ile Tyr Met Val Pro Gln Asn Ile Lys Pro Ala Leu Gln Arg Thr Ala |                 |            |            |             |             |     |
|   | 230 235 240     |            |            |             |             |     |

64/68

|  |      |
|--|------|
| atc gag atc ctg gca tgg ggc ctg cgg aac atg aag agt tac cag ctg<br>Ile Glu Ile Leu Ala Trp Gly Leu Arg Asn Met Lys Ser Tyr Gln Leu | 1188 |
| 245 250 255  |      |
| gcc aac atc tcc tcc ccc agc ctc gtg gta gag tgt ggg ggc cag acg<br>Ala Asn Ile Ser Ser Pro Ser Leu Val Val Glu Cys Gly Gly Gln Thr | 1236 |
| 260 265 270  |      |
| gtg cag tcc tgt gtc atc agg aac ctc cgg aag aac ccc aac ttt gac<br>Val Gln Ser Cys Val Ile Arg Asn Leu Arg Lys Asn Pro Asn Phe Asp | 1284 |
| 275 280 285  |      |
| atc tgc acc ctc ttc atg gaa gtg atg ctg ccc agg gag gag ctc tac<br>Ile Cys Thr Leu Phe Met Glu Val Met Leu Pro Arg Glu Glu Leu Tyr | 1332 |
| 290 295 300 305  |      |
| tgc ccc ccc atc acc gtc aag gtc atc gat aac cgc cag ttt ggc cgc<br>Cys Pro Pro Ile Thr Val Lys Val Ile Asp Asn Arg Gln Phe Gly Arg | 1380 |
| 310 315 320  |      |
| cgg cct gtg gtg ggc cag tgt acc atc cgc tcc ctg gag agc ttc ctg<br>Arg Pro Val Val Gly Gln Cys Thr Ile Arg Ser Leu Glu Ser Phe Leu | 1428 |
| 325 330 335  |      |
| tgt gac ccc tac tcg gcg gag agt cca tcc cca cag ggt ggc cca gac<br>Cys Asp Pro Tyr Ser Ala Glu Ser Pro Ser Pro Gln Gly Gly Pro Asp | 1476 |
| 340 345 350  |      |
| gat gtg agc cta ctc agt cct ggg gaa gac gtg ctc atc gac att gat<br>Asp Val Ser Leu Leu Ser Pro Gly Glu Asp Val Leu Ile Asp Ile Asp | 1524 |
| 355 360 365  |      |
| gac aag gag ccc ctc atc ccc atc cag gag gaa gag ttc atc gat tgg<br>Asp Lys Glu Pro Leu Ile Pro Ile Gln Glu Glu Phe Ile Asp Trp     | 1572 |
| 370 375 380 385  |      |
| tgg agc aaa ttc ttt gcc tcc ata ggg gag agg gaa aag tgc ggc tcc<br>Trp Ser Lys Phe Phe Ala Ser Ile Gly Glu Arg Glu Lys Cys Gly Ser | 1620 |
| 390 395 400  |      |
| tac ctg gag aag gat ttt gac acc ctg aag gtc tat gac aca cag ctg<br>Tyr Leu Glu Lys Asp Phe Asp Thr Leu Lys Val Tyr Asp Thr Gln Leu | 1668 |
| 405 410 415  |      |
| gag aat gtg gag gcc ttt gag ggc ctg tct gac ttt tgt aac acc ttc<br>Glu Asn Val Glu Ala Phe Glu Gly Leu Ser Asp Phe Cys Asn Thr Phe | 1716 |
| 420 425 430  |      |
| aag ctg tac cgg ggc aag acg cag gag gag aca gaa gat cca tct gtg<br>Lys Leu Tyr Arg Gly Lys Thr Gln Glu Glu Thr Glu Asp Pro Ser Val | 1764 |
| 435 440 445  |      |
| att ggt gaa ttt aag ggc ctc ttc aaa att tat ccc ctc cca gaa gac<br>Ile Gly Glu Phe Lys Gly Leu Phe Lys Ile Tyr Pro Leu Pro Glu Asp | 1812 |
| 450 455 460 465  |      |
| cca gcc atc ccc atg ccc cca aga cag ttc cac cag ctg gcc gcc cag<br>Pro Ala Ile Pro Met Pro Pro Arg Gln Phe His Gln Leu Ala Ala Gln | 1860 |
| 470 475 480  |      |
| gga ccc cag gag tgc ttg gtc cgt atc tac att gtc cga gca ttt ggc<br>Gly Pro Gln Glu Cys Leu Val Arg Ile Tyr Ile Val Arg Ala Phe Gly | 1908 |
| 485 490 495  |      |
| ctg cag ccc aag gac ccc aat gga aag tgt gat cct tac atc aag atc<br>Leu Gln Pro Lys Asp Pro Asn Gly Lys Cys Asp Pro Tyr Ile Lys Ile | 1956 |
| 500 505 510  |      |

65/68

|                   |            |            |            |            |            |            |            |            |            |            |            |            |            |            |            |      |
|-------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------|
| tcc<br>Ser<br>515 | ata<br>Ile | ggg<br>Gly | aag<br>Lys | aaa<br>Lys | tca<br>Ser | gtg<br>Val | agt<br>Ser | gac<br>Asp | cag<br>Gln | gat<br>Asp | aac<br>Asn | tac<br>Tyr | atc<br>Ile | ccc<br>Pro | tgc<br>Cys | 2004 |
| acg<br>Thr<br>530 | ctg<br>Leu | gag<br>Glu | ccc<br>Pro | gta<br>Val | ttt<br>Phe | gga<br>Gly | aag<br>Lys | atg<br>Met | ttc<br>Phe | gag<br>Glu | ctg<br>Leu | acc<br>Thr | tgc<br>Cys | act<br>Thr | ctg<br>Leu | 2052 |
| cct<br>Pro        | ctg<br>Leu | gag<br>Glu | aag<br>Lys | gac<br>Asp | cta<br>Leu | aag<br>Lys | atc<br>Ile | act<br>Thr | ctc<br>Leu | tat<br>Tyr | gac<br>Asp | tat<br>Tyr | gac<br>Asp | ctc<br>Leu | ctc<br>Leu | 2100 |
| tcc<br>Ser        | aag<br>Lys | gac<br>Asp | gaa<br>Glu | aag<br>Lys | atc<br>Ile | ggt<br>Gly | gag<br>Glu | acg<br>Thr | gtc<br>Val | gtc<br>Val | gac<br>Asp | ctg<br>Leu | gag<br>Glu | aac<br>Asn | agg<br>Arg | 2148 |
| ctg<br>Leu        | ctg<br>Leu | tcc<br>Ser | aag<br>Lys | ttt<br>Phe | ggg<br>Gly | gct<br>Ala | cgc<br>Arg | tgt<br>Cys | gga<br>Gly | ctc<br>Leu | cca<br>Pro | cag<br>Gln | acc<br>Thr | tac<br>Tyr | tgt<br>Cys | 2196 |
| gtc<br>Val        | tct<br>Ser | gga<br>Gly | ccg<br>Pro | aac<br>Asn | cag<br>Gln | tgg<br>Trp | cgg<br>Arg | gac<br>Asp | cag<br>Gln | ctc<br>Leu | cgc<br>Arg | ccc<br>Pro | tcc<br>Ser | cag<br>Gln | ctc<br>Leu | 2244 |
| ctc<br>Leu        | cac<br>His | ctc<br>Leu | ttc<br>Phe | tgc<br>Cys | cag<br>Gln | cag<br>Gln | cat<br>His | aga<br>Arg | gtc<br>Val | aag<br>Lys | gca<br>Ala | cct<br>Pro | gtg<br>Val | tac<br>Tyr | cgg<br>Arg | 2292 |
| aca<br>Thr        | gac<br>Asp | cgt<br>Arg | gta<br>Val | atg<br>Met | ttt<br>Phe | cag<br>Gln | gat<br>Asp | aaa<br>Lys | gaa<br>Glu | tat<br>Tyr | tcc<br>Ser | att<br>Ile | gaa<br>Glu | gag<br>Glu | ata<br>Ile | 2340 |
| gag<br>Glu        | gct<br>Ala | ggc<br>Gly | agg<br>Arg | atc<br>Ile | cca<br>Pro | aac<br>Asn | cca<br>Pro | cac<br>His | ctg<br>Leu | ggc<br>Gly | cca<br>Pro | gtg<br>Val | gag<br>Glu | gag<br>Glu | cgt<br>Arg | 2388 |
| ctg<br>Leu        | gct<br>Ala | ctg<br>Leu | cat<br>His | gtg<br>Val | ctt<br>Leu | cag<br>Gln | cag<br>Gln | cag<br>Gln | ggc<br>Gly | ctg<br>Leu | gtc<br>Val | ccg<br>Pro | gag<br>Glu | cac<br>His | gtg<br>Val | 2436 |
| gag<br>Glu        | tca<br>Ser | cgg<br>Arg | ccc<br>Pro | ctc<br>Leu | tac<br>Tyr | agc<br>Ser | ccc<br>Pro | ctg<br>Leu | cag<br>Gln | cca<br>Pro | gac<br>Asp | atc<br>Ile | gag<br>Glu | cag<br>Gln | ggg<br>Gly | 2484 |
| aag<br>Lys        | ctg<br>Leu | cag<br>Gln | atg<br>Met | tgg<br>Trp | gtc<br>Val | gac<br>Asp | cta<br>Leu | ttt<br>Phe | ccg<br>Pro | aag<br>Lys | gcc<br>Ala | ctg<br>Leu | ggg<br>Gly | cgg<br>Arg | cct<br>Pro | 2532 |
| gga<br>Gly        | cct<br>Pro | ccc<br>Pro | ttc<br>Phe | aac<br>Asn | atc<br>Ile | acc<br>Thr | cca<br>Pro | cgg<br>Arg | aga<br>Arg | gcc<br>Ala | aga<br>Arg | agg<br>Arg | ttt<br>Phe | ttc<br>Phe | ctg<br>Leu | 2580 |
| cgt<br>Arg        | tgt<br>Cys | att<br>Ile | atc<br>Ile | tgg<br>Trp | aat<br>Asn | acc<br>Thr | aga<br>Arg | gat<br>Asp | gtg<br>Val | atc<br>Ile | ctg<br>Leu | gat<br>Asp | gac<br>Asp | ctg<br>Leu | agc<br>Ser | 2628 |
| ctc<br>Leu        | acg<br>Thr | ggg<br>Gly | gag<br>Glu | aag<br>Lys | atg<br>Met | agc<br>Ser | gac<br>Asp | att<br>Ile | tat<br>Tyr | gtg<br>Val | aaa<br>Lys | ggt<br>Gly | tgg<br>Trp | atg<br>Met | att<br>Ile | 2676 |
| ggc<br>Gly        | ttt<br>Phe | gaa<br>Glu | gaa<br>Glu | cac<br>His | aag<br>Lys | caa<br>Gln | aag<br>Lys | aca<br>Thr | gac<br>Asp | gtg<br>Val | cat<br>His | tat<br>Tyr | cgt<br>Arg | tcc<br>Ser | ctg<br>Leu | 2724 |



66/68

|  |      |
|--|------|
| gga ggt gaa ggc aac ttc aac tgg agg ttc att ttc ccc ttc gac tac    | 2772 |
| Gly Gly Glu Gly Asn Phe Asn Trp Arg Phe Ile Phe Pro Phe Asp Tyr    |      |
| 770 775 780 785  |      |
| ctg cca gct gag caa gtc tgt acc att gcc aag aag gat gcc ttc tgg    | 2820 |
| Leu Pro Ala Glu Val Cys Thr Ile Ala Lys Lys Asp Ala Phe Trp        |      |
| 790 795 800  |      |
| agg ctg gac aag act gag agc aaa atc cca gca cga gtg gtg ttc cag    | 2868 |
| Arg Leu Asp Lys Thr Glu Ser Lys Ile Pro Ala Arg Val Val Phe Gln    |      |
| 805 810 815  |      |
| atc tgg gac aat gac aag ttc tcc ttt gat gat ttt ctg ggc tcc ctg    | 2916 |
| Ile Trp Asp Asn Asp Lys Phe Ser Phe Asp Asp Phe Leu Gly Ser Leu    |      |
| 820 825 830  |      |
| cag ctc gat ctc aac cgc atg ccc aag cca gcc aag aca gcc aag aag    | 2964 |
| Gln Leu Asp Leu Asn Arg Met Pro Lys Pro Ala Lys Thr Ala Lys Lys    |      |
| 835 840 845  |      |
| tgc tcc ttg gac cag ctg gat gat gct ttc cac cca gaa tgg ttt gtg    | 3012 |
| Cys Ser Leu Asp Gln Leu Asp Asp Ala Phe His Pro Glu Trp Phe Val    |      |
| 850 855 860 865  |      |
| tcc ctt ttt gag cag aaa aca gtg aag ggc tgg tgg ccc tgt gta gca    | 3060 |
| Ser Leu Phe Glu Gln Lys Thr Val Lys Gly Trp Trp Pro Cys Val Ala    |      |
| 870 875 880  |      |
| gaa gag ggt gag aag aaa ata ctg gcg ggc aag ctg gaa atg acc ttg    | 3108 |
| Glu Glu Gly Glu Lys Lys Ile Leu Ala Gly Lys Leu Glu Met Thr Leu    |      |
| 885 890 895  |      |
| gag att gta gca gag agt gag cat gag gag cgg cct gct ggc cag ggc    | 3156 |
| Glu Ile Val Ala Glu Ser Glu His Glu Glu Arg Pro Ala Gly Gln Gly    |      |
| 900 905 910  |      |
| cgg gat gag ccc aac atg aac cct aag ctt gag gac cca agg cgc ccc    | 3204 |
| Arg Asp Glu Pro Asn Met Asn Pro Lys Leu Glu Asp Pro Arg Arg Pro    |      |
| 915 920 925  |      |
| gac acc tcc ttc ctg tgg ttt acc tcc cca tac aag acc atg aag ttc    | 3252 |
| Asp Thr Ser Phe Leu Trp Phe Thr Ser Pro Tyr Lys Thr Met Lys Phe    |      |
| 930 935 940 945  |      |
| atc ctg tgg cgg cgt ttc cgg tgg gcc atc atc ctc ttc atc atc ctc    | 3300 |
| Ile Leu Trp Arg Arg Phe Arg Trp Ala Ile Ile Leu Phe Ile Ile Leu    |      |
| 950 955 960  |      |
| ttc atc ctg ctg ctg ttc ctg gcc atc ttc atc tac gcc ttc ccg aac    | 3348 |
| Phe Ile Leu Leu Leu Phe Leu Ala Ile Phe Ile Tyr Ala Phe Pro Asn    |      |
| 965 970 975  |      |
| tat gct gcc atg aag ctg gtg aag ccc ttc agc tgaggactct cctgcctgt   | 3401 |
| Tyr Ala Ala Met Lys Leu Val Lys Pro Phe Ser                        |      |
| 980 985  |      |
| agaagggggcc gtgggggtccc ctccagcatg ggactggcct gcctcctccg ccagctcgg | 3461 |
| cgagctcctc cagacctcct aggctgatt gtccctgccag ggtgggcaga cagacagatg  | 3521 |
| gaccggccca cactcccaga gttgctaaca tggagctctg agatcacccc acttccatca  | 3581 |
| tttcttctc ccccaaccca acgctttttt ggatcagctc agacatatatt cagtataaaa  | 3641 |
| cagttggaac cacaaaaaaaa aaaaaaaaaa                                  | 3671 |

&lt;210&gt; 233

&lt;211&gt; 988

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

67/68

<400> 233  
 Met Leu Gly Arg Leu Ser Leu Glu Lys Gly Arg Phe Val Asn Pro Gly  
 1 5 10 15  
 Gly Arg Gly Arg Asp Pro Gly Glu Gly Gly Val Met Asp Asp Lys Ser  
 20 25 30  
 Glu Asp Ser Met Ser Val Ser Thr Leu Ser Phe Gly Val Asn Arg Pro  
 35 40 45  
 Thr Ile Ser Cys Ile Phe Asp Tyr Gly Asn Arg Tyr His Leu Arg Cys  
 50 55 60  
 Tyr Met Tyr Gln Ala Arg Asp Leu Ala Ala Met Asp Lys Asp Ser Phe  
 65 70 75 80  
 Ser Asp Pro Tyr Ala Ile Val Ser Phe Leu His Gln Ser Gln Lys Thr  
 85 90 95  
 Val Val Val Lys Asn Thr Leu Asn Pro Thr Trp Asp Gln Thr Leu Ile  
 100 105 110  
 Phe Tyr Glu Ile Glu Ile Phe Gly Glu Pro Ala Thr Val Ala Glu Gln  
 115 120 125  
 Pro Pro Ser Ile Val Val Glu Leu Tyr Asp His Asp Thr Tyr Gly Ala  
 130 135 140  
 Asp Glu Phe Met Gly Arg Cys Ile Cys Gln Pro Ser Leu Glu Arg Met  
 145 150 155 160  
 Pro Arg Leu Ala Trp Phe Pro Leu Thr Arg Gly Ser Gln Pro Ser Gly  
 165 170 175  
 Glu Leu Leu Ala Ser Phe Glu Leu Ile Gln Arg Glu Lys Pro Ala Ile  
 180 185 190  
 His His Ile Pro Gly Phe Glu Val Gln Glu Thr Ser Arg Ile Leu Asp  
 195 200 205  
 Glu Ser Glu Asp Thr Asp Leu Pro Tyr Pro Pro Pro Gln Arg Glu Ala  
 210 215 220  
 Asn Ile Tyr Met Val Pro Gln Asn Ile Lys Pro Ala Leu Gln Arg Thr  
 225 230 235 240  
 Ala Ile Glu Ile Leu Ala Trp Gly Leu Arg Asn Met Lys Ser Tyr Gln  
 245 250 255  
 Leu Ala Asn Ile Ser Ser Pro Ser Leu Val Val Glu Cys Gly Gly Gln  
 260 265 270  
 Thr Val Gln Ser Cys Val Ile Arg Asn Leu Arg Lys Asn Pro Asn Phe  
 275 280 285  
 Asp Ile Cys Thr Leu Phe Met Glu Val Met Leu Pro Arg Glu Glu Leu  
 290 295 300  
 Tyr Cys Pro Pro Ile Thr Val Lys Val Ile Asp Asn Arg Gln Phe Gly  
 305 310 315 320  
 Arg Arg Pro Val Val Gly Gln Cys Thr Ile Arg Ser Leu Glu Ser Phe  
 325 330 335  
 Leu Cys Asp Pro Tyr Ser Ala Glu Ser Pro Ser Pro Gln Gly Gly Pro  
 340 345 350  
 Asp Asp Val Ser Leu Leu Ser Pro Gly Glu Asp Val Leu Ile Asp Ile  
 355 360 365  
 Asp Asp Lys Glu Pro Leu Ile Pro Ile Gln Glu Glu Glu Phe Ile Asp  
 370 375 380  
 Trp Trp Ser Lys Phe Phe Ala Ser Ile Gly Glu Arg Glu Lys Cys Gly  
 385 390 395 400  
 Ser Tyr Leu Glu Lys Asp Phe Asp Thr Leu Lys Val Tyr Asp Thr Gln  
 405 410 415  
 Leu Glu Asn Val Glu Ala Phe Glu Gly Leu Ser Asp Phe Cys Asn Thr  
 420 425 430  
 Phe Lys Leu Tyr Arg Gly Lys Thr Gln Glu Glu Thr Glu Asp Pro Ser  
 435 440 445  
 Val Ile Gly Glu Phe Lys Gly Leu Phe Lys Ile Tyr Pro Leu Pro Glu  
 450 455 460  
 Asp Pro Ala Ile Pro Met Pro Pro Arg Gln Phe His Gln Leu Ala Ala  
 465 470 475 480  
 Gln Gly Pro Gln Glu Cys Leu Val Arg Ile Tyr Ile Val Arg Ala Phe  
 485 490 495  
 Gly Leu Gln Pro Lys Asp Pro Asn Gly Lys Cys Asp Pro Tyr Ile Lys  
 500 505 510  
 Ile Ser Ile Gly Lys Lys Ser Val Ser Asp Gln Asp Asn Tyr Ile Pro  
 515 520 525

68/68

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Cys | Thr | Leu | Glu | Pro | Val | Phe | Gly | Lys | Met | Phe | Glu | Leu | Thr | Cys | Thr |
| 530 |     |     |     |     |     | 535 |     |     |     |     | 540 |     |     |     |     |
| Leu | Pro | Leu | Glu | Lys | Asp | Leu | Lys | Ile | Thr | Leu | Tyr | Asp | Tyr | Asp | Leu |
| 545 |     |     |     |     | 550 |     |     |     |     | 555 |     |     |     |     | 560 |
| Leu | Ser | Lys | Asp | Glu | Lys | Ile | Gly | Glu | Thr | Val | Val | Asp | Leu | Glu | Asn |
|     |     |     |     | 565 |     |     |     |     | 570 |     |     |     |     | 575 |     |
| Arg | Leu | Leu | Ser | Lys | Phe | Gly | Ala | Arg | Cys | Gly | Leu | Pro | Gln | Thr | Tyr |
|     |     |     | 580 |     |     |     |     | 585 |     |     |     |     | 590 |     |     |
| Cys | Val | Ser | Gly | Pro | Asn | Gln | Trp | Arg | Asp | Gln | Leu | Arg | Pro | Ser | Gln |
|     |     | 595 |     |     |     |     | 600 |     |     |     |     | 605 |     |     |     |
| Leu | Leu | His | Leu | Phe | Cys | Gln | Gln | His | Arg | Val | Lys | Ala | Pro | Val | Tyr |
|     | 610 |     |     |     |     | 615 |     |     |     |     | 620 |     |     |     |     |
| Arg | Thr | Asp | Arg | Val | Met | Phe | Gln | Asp | Lys | Glu | Tyr | Ser | Ile | Glu | Glu |
| 625 |     |     |     |     | 630 |     |     |     |     | 635 |     |     |     |     | 640 |
| Ile | Glu | Ala | Gly | Arg | Ile | Pro | Asn | Pro | His | Leu | Gly | Pro | Val | Glu | Glu |
|     |     |     |     | 645 |     |     |     |     | 650 |     |     |     |     | 655 |     |
| Arg | Leu | Ala | Leu | His | Val | Leu | Gln | Gln | Gln | Gly | Leu | Val | Pro | Glu | His |
|     |     |     | 660 |     |     |     |     | 665 |     |     |     |     | 670 |     |     |
| Val | Glu | Ser | Arg | Pro | Leu | Tyr | Ser | Pro | Leu | Gln | Pro | Asp | Ile | Glu | Gln |
|     |     | 675 |     |     |     |     | 680 |     |     |     |     | 685 |     |     |     |
| Gly | Lys | Leu | Gln | Met | Trp | Val | Asp | Leu | Phe | Pro | Lys | Ala | Leu | Gly | Arg |
|     | 690 |     |     |     |     | 695 |     |     |     |     | 700 |     |     |     |     |
| Pro | Gly | Pro | Pro | Phe | Asn | Ile | Thr | Pro | Arg | Arg | Ala | Arg | Arg | Phe | Phe |
| 705 |     |     |     |     | 710 |     |     |     |     | 715 |     |     |     |     | 720 |
| Leu | Arg | Cys | Ile | Ile | Trp | Asn | Thr | Arg | Asp | Val | Ile | Leu | Asp | Asp | Leu |
|     |     |     |     | 725 |     |     |     |     | 730 |     |     |     |     | 735 |     |
| Ser | Leu | Thr | Gly | Glu | Lys | Met | Ser | Asp | Ile | Tyr | Val | Lys | Gly | Trp | Met |
|     |     |     | 740 |     |     |     |     | 745 |     |     |     |     | 750 |     |     |
| Ile | Gly | Phe | Glu | Glu | His | Lys | Gln | Lys | Thr | Asp | Val | His | Tyr | Arg | Ser |
|     |     | 755 |     |     |     |     | 760 |     |     |     |     | 765 |     |     |     |
| Leu | Gly | Gly | Glu | Gly | Asn | Phe | Asn | Trp | Arg | Phe | Ile | Phe | Pro | Phe | Asp |
|     | 770 |     |     |     |     | 775 |     |     |     |     | 780 |     |     |     |     |
| Tyr | Leu | Pro | Ala | Glu | Gln | Val | Cys | Thr | Ile | Ala | Lys | Lys | Asp | Ala | Phe |
| 785 |     |     |     |     | 790 |     |     |     |     | 795 |     |     |     |     | 800 |
| Trp | Arg | Leu | Asp | Lys | Thr | Glu | Ser | Lys | Ile | Pro | Ala | Arg | Val | Val | Phe |
|     |     |     |     | 805 |     |     |     |     | 810 |     |     |     |     | 815 |     |
| Gln | Ile | Trp | Asp | Asn | Asp | Lys | Phe | Ser | Phe | Asp | Asp | Phe | Leu | Gly | Ser |
|     |     |     | 820 |     |     |     |     | 825 |     |     |     |     | 830 |     |     |
| Leu | Gln | Leu | Asp | Leu | Asn | Arg | Met | Pro | Lys | Pro | Ala | Lys | Thr | Ala | Lys |
|     |     | 835 |     |     |     |     | 840 |     |     |     |     | 845 |     |     |     |
| Lys | Cys | Ser | Leu | Asp | Gln | Leu | Asp | Asp | Ala | Phe | His | Pro | Glu | Trp | Phe |
|     | 850 |     |     |     |     | 855 |     |     |     |     | 860 |     |     |     |     |
| Val | Ser | Leu | Phe | Glu | Gln | Lys | Thr | Val | Lys | Gly | Trp | Trp | Pro | Cys | Val |
| 865 |     |     |     |     | 870 |     |     |     |     | 875 |     |     |     |     | 880 |
| Ala | Glu | Glu | Gly | Glu | Lys | Lys | Ile | Leu | Ala | Gly | Lys | Leu | Glu | Met | Thr |
|     |     |     |     | 885 |     |     |     |     | 890 |     |     |     |     | 895 |     |
| Leu | Glu | Ile | Val | Ala | Glu | Ser | Glu | His | Glu | Glu | Arg | Pro | Ala | Gly | Gln |
|     |     |     | 900 |     |     |     |     | 905 |     |     |     |     | 910 |     |     |
| Gly | Arg | Asp | Glu | Pro | Asn | Met | Asn | Pro | Lys | Leu | Glu | Asp | Pro | Arg | Arg |
|     |     | 915 |     |     |     |     | 920 |     |     |     |     | 925 |     |     |     |
| Pro | Asp | Thr | Ser | Phe | Leu | Trp | Phe | Thr | Ser | Pro | Tyr | Lys | Thr | Met | Lys |
|     | 930 |     |     |     |     | 935 |     |     |     |     | 940 |     |     |     |     |
| Phe | Ile | Leu | Trp | Arg | Arg | Phe | Arg | Trp | Ala | Ile | Ile | Leu | Phe | Ile | Ile |
| 945 |     |     |     |     | 950 |     |     |     |     | 955 |     |     |     |     | 960 |
| Leu | Phe | Ile | Leu | Leu | Leu | Phe | Leu | Ala | Ile | Phe | Ile | Tyr | Ala | Phe | Pro |
|     |     |     |     | 965 |     |     |     |     | 970 |     |     |     |     | 975 |     |
| Asn | Tyr | Ala | Ala | Met | Lys | Leu | Val | Lys | Pro | Phe | Ser |     |     |     |     |
|     |     |     | 980 |     |     |     |     | 985 |     |     |     |     |     |     |     |

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US99/19395**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(7) : C12N 15/11, 15/00; C07K 16/00

US CL : 536/23.1, 435/440, 530/387.1

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1, 435/440, 530/387.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

BIOSIS, CAPLUS, EMBASE, EMBASE, EMBASE, LIFESCI, MEDLINE, SCISEARCH, TOXLIT

Search Terms: dysferlin, lgmd2b

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages  | Relevant to claim No. |
|-----------|---|-----------------------|
| X         | WEILER et al. Limb-girdle muscular dystrophy and Myoshi Myopathy in an aboriginal Canadian kindred map to LGMD2B and segregate with the same haplotype. American Journal of Human Genetics. October 1996, Vol.59, pages 872-878, especially page 873. | 32,35                 |
| X         | KOENIG et al. Complete cloning of the Duchenne Muscular Dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell. 31 July 1987, Vol. 50, pages 509-517, especially pages 511-513.           | 32-33,36              |

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

|   |  |
|---|--|
| * Special categories of cited documents:  | *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention  |
| *A* document defining the general state of the art which is not considered to be of particular relevance  | *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone   |
| *B* earlier document published on or after the international filing date  | *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | *Z* document member of the same patent family  |
| *O* document referring to an oral disclosure, use, exhibition or other means  |  |
| *P* document published prior to the international filing date but later than the priority date claimed  |  |

|   |  |
|---|--|
| Date of the actual completion of the international search<br>17 NOVEMBER 1999   | Date of mailing of the international search report<br><b>13 JAN 2000</b> |
| Name and mailing address of the ISA/US<br>Commissioner of Patents and Trademarks<br>Box PCT<br>Washington, D.C. 20231<br>Facsimile No. (703) 305-4242 | Authorized officer<br>Stephen Siu<br>Telephone No. (703) 308-0196        |

Form PCT/ISA/210 (second sheet)(July 1992)\*

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US99/19395

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category*         | Citation of document, with indication, where appropriate, of the relevant passages   | Relevant to claim No.         |
|-------------------|--|-------------------------------|
| X,P<br>---<br>Y,P | Database GenCore version 4.5, Compugen Ltd., No. AI128455, 'NCI-CGAP, National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index', Unpublished, 27 October 1998   | 1,6,12<br>-----<br>7,14,16    |
| X<br>---<br>Y     | Database GenCore version 4.5, Compugen Ltd., No. R41062, WAYE, M.M.Y. et al. 'Gene expression of adult human heart as revealed by random sequencing of cDNA library,' Miami Winter Biotechnol. Symp. Proc. 6,90 , 16 May 16, 1995. | 1, 6, 11-12<br>-----<br>7, 14 |
| X<br>---<br>Y     | Database GenCore version 4.5, Compugen Ltd., No. AA718275, Marra et al, 'The WashU-HHMI Mouse EST Project', Unpublished, 29 December 1997.   | 1, 6, 11-12<br>-----<br>7, 14 |
| Y                 | BASHIR et al. Genetic and physical mapping at the limb-girdle muscular dystrophy locus (LGMD2B) on chromosome 2p. Genomics. April 1996, Vol.33, pages 46-52, especially page 47.   | 32,36                         |
| Y                 | MOREIRA et al. The seventh form of autosomal recessive limb-girdle muscular dystrophy is mapped to 17q11-12. American Journal of Human Genetics. July 1997, Vol. 61, pages 151-159, entire document.                               | 32, 35                        |
| Y                 | Database GenCore version 4.5, Compugen Ltd., No. R76778, HILLIER et al., 'The WashU-Merck EST Project', Unpublished, 06 June 1995.   | 7, 14                         |
| A,E               | AHLBERG et al. Genetic Linkage of Welander Distal Myopathy to chromosome 2p13. Annals of Neurology. September 1999, Vol. 46, No.3, pages 399-404, especially page 400.   | 37, 39                        |
| A,E               | BITTNER et al. Dysferlin deletion in SJL mice (SJL-Dysf) defines a natural model for limb girdle muscular dystrophy 2B. Nature Genetics. October 1999, Vol. 23, pages 141-142, especially page 141.                                | 40                            |
| A,P               | BASHIR et al. A gene related to Caenorhabditis elegans spermatogenesis factor fer-1 is mutated in limb-girdle muscular dystrophy type 2B. Nature Genetics. September 1998, Vol 20, pages 37-42.                                    | 1-53                          |
| A,E               | Matsuda et al. Dysferlin is a surface membrane-associated protein that is absent in Miyoshi Myopathy. Neurology 22 September 1999, Vol. 53, No. 5, pages 1119-1122, especially pages 1119-1120.                                    | 40                            |

Form PCT/ISA/210 (continuation of second sheet)(July 1992)\*

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US99/19395

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

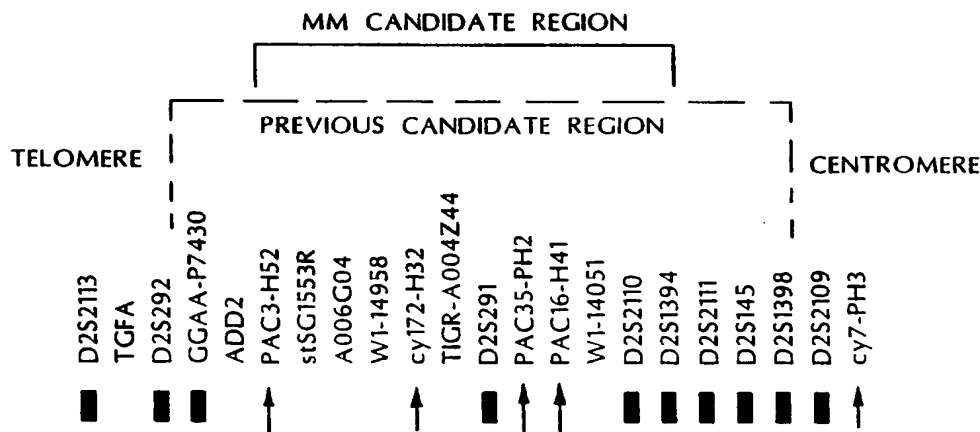
| Category* | Citation of document, with indication, where appropriate, of the relevant passages  | Relevant to claim No. |
|-----------|---|-----------------------|
| A,P       | LIU et al. Dysferlin, a novel skeletal muscle gene, is mutated in Miyoshi Myopathy and limb girdle muscular dystrophy. Nature Genetics. September 1998, Vol. 20, pages 31-36. | 1-54                  |

Form PCT/ISA/210 (continuation of second sheet)(July 1992)★



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

|  |  |           |  |
|--|--|-----------|--|
| (51) International Patent Classification <sup>7</sup> :<br><b>C12N 15/11, 15/00, C07K 16/00</b>  |  | <b>A1</b> | (11) International Publication Number: <b>WO 00/11157</b>  |
|  |  |           | (43) International Publication Date: <b>2 March 2000 (02.03.00)</b>  |
| (21) International Application Number: <b>PCT/US99/19395</b>   |  |           | (81) Designated States: CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).<br><br><b>Published</b><br><i>With international search report.<br/>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i> |
| (22) International Filing Date: <b>25 August 1999 (25.08.99)</b>   |  |           |  |
| (30) Priority Data:<br><b>60/097,927</b> <b>25 August 1998 (25.08.98)</b> <b>US</b>  |  |           |  |
| (71) Applicant: <b>THE GENERAL HOSPITAL CORPORATION</b><br>[US/US]; 55 Fruit Street, Boston, MA 02114 (US).  |  |           |  |
| (72) Inventors: <b>BROWN, Robert, H.</b> ; 16 Oakland Avenue, Needham, MA 02192 (US). <b>LIU, Jing</b> ; 1629 Van Horn, Outremont, Quebec H2V 1L1 (CA). <b>AOKI, Masashi</b> ; Dept. of Neurology, Tohoku University School of Medicine, 1-1, Seiro-Machi, Aoba-ku, Sendai (JP). <b>HO, Meng, F.</b> ; Apartment 24, 145 Englewood Avenue, Brighton, MA 02135 (US). <b>MATSUDA-ASADA, Chie</b> ; 33 Pond Avenue, Brookline, MA 02445 (US). |  |           |  |
| (74) Agent: <b>FRASER, Janis, K.</b> ; Fish & Richardson P.C., 225 Franklin Street, Boston, MA 02110-2804 (US).  |  |           |  |

(54) Title: **DYSFERLIN, A GENE MUTATED IN DISTAL MYOPATHY AND LIMB GIRDLE MUSCULAR DYSTROPHY**

## (57) Abstract

A novel gene and the protein encoded therein, i.e., dysferlin, are disclosed. This gene and its expression products are associated with muscular dystrophy, e.g., Miyoshi myopathy and limb girdle muscular dystrophy 2B.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

|    |                          |    |  |    |  |    |                          |
|----|--------------------------|----|--|----|--|----|--------------------------|
| AL | Albania                  | ES | Spain                                    | LS | Lesotho                                      | SI | Slovenia                 |
| AM | Armenia                  | FI | Finland                                  | LT | Lithuania                                    | SK | Slovakia                 |
| AT | Austria                  | FR | France                                   | LU | Luxembourg                                   | SN | Senegal                  |
| AU | Australia                | GA | Gabon                                    | LV | Latvia                                       | SZ | Swaziland                |
| AZ | Azerbaijan               | GB | United Kingdom                           | MC | Monaco                                       | TD | Chad                     |
| BA | Bosnia and Herzegovina   | GE | Georgia                                  | MD | Republic of Moldova                          | TG | Togo                     |
| BB | Barbados                 | GH | Ghana                                    | MG | Madagascar                                   | TJ | Tajikistan               |
| BE | Belgium                  | GN | Guinea                                   | MK | The former Yugoslav<br>Republic of Macedonia | TM | Turkmenistan             |
| BF | Burkina Faso             | GR | Greece                                   |    |  | TR | Turkey                   |
| BG | Bulgaria                 | HU | Hungary                                  | ML | Mali   | TT | Trinidad and Tobago      |
| BJ | Benin                    | IE | Ireland                                  | MN | Mongolia                                     | UA | Ukraine                  |
| BR | Brazil                   | IL | Israel                                   | MR | Mauritania                                   | UG | Uganda                   |
| BY | Belarus                  | IS | Iceland                                  | MW | Malawi                                       | US | United States of America |
| CA | Canada                   | IT | Italy                                    | MX | Mexico                                       | UZ | Uzbekistan               |
| CF | Central African Republic | JP | Japan                                    | NE | Niger  | VN | Viet Nam                 |
| CG | Congo                    | KE | Kenya                                    | NL | Netherlands                                  | YU | Yugoslavia               |
| CH | Switzerland              | KG | Kyrgyzstan                               | NO | Norway                                       | ZW | Zimbabwe                 |
| CI | Côte d'Ivoire            | KP | Democratic People's<br>Republic of Korea | NZ | New Zealand                                  |    |                          |
| CM | Cameroon                 | KR | Republic of Korea                        | PL | Poland                                       |    |                          |
| CN | China                    | KZ | Kazakhstan                               | PT | Portugal                                     |    |                          |
| CU | Cuba                     | LC | Saint Lucia                              | RO | Romania                                      |    |                          |
| CZ | Czech Republic           | LI | Liechtenstein                            | RU | Russian Federation                           |    |                          |
| DE | Germany                  | LK | Sri Lanka                                | SD | Sudan  |    |                          |
| DK | Denmark                  | LR | Liberia                                  | SE | Sweden                                       |    |                          |
| EE | Estonia                  |    |  | SG | Singapore                                    |    |                          |



DYSFERLIN, A GENE MUTATED IN DISTAL MYOPATHY  
AND LIMB GIRDLE MUSCULAR DYSTROPHY

5                   RELATED APPLICATION INFORMATION

This application claims priority from provisional application serial no. 60/097,927, filed August 25, 1998.

Statement as to Federally Sponsored Research

The work described herein was supported in part by  
10 NIH grants 5P01AG12992, 5R01N834913A, and 5P01NS31248.  
The Federal Government therefore may have certain rights  
in the invention.

Background of the Invention

The invention relates to genes involved in the  
15 onset of muscular dystrophy.

Muscular dystrophies constitute a heterogeneous group of disorders. Most are characterized by weakness and atrophy of the proximal muscles, although in rare myopathies such as "Miyoshi myopathy" symptoms may first  
20 arise in distal muscles. Of the various hereditary types of muscular dystrophy, several are caused by mutations or deletions in genes encoding individual components of the dystrophin-associated protein (DAP) complex. It is this DAP complex that links the cytoskeletal protein  
25 dystrophin to the extracellular matrix protein, laminin-2.

Muscular dystrophies may be classified according to the gene mutations that are associated with specific clinical syndromes. For example, mutations in the gene  
30 encoding the cytoskeletal protein dystrophin result in either Duchenne's Muscular Dystrophy or Becker's Muscular Dystrophy, whereas mutations in the gene encoding the extracellular matrix protein merosin produce Congenital

- 2 -

Muscular Dystrophy. Muscular dystrophies with an autosomal recessive mode of inheritance include "Miyoshi myopathy" and the several limb-girdle muscular dystrophies (LGMD2). Of the limb-girdle muscular dystrophies, the deficiencies resulting in LGMD2C, D, E, and F result from mutations in genes encoding the membrane-associated sarcoglycan components of the DAP complex.

#### Summary of the Invention

10 A novel protein, designated dysferlin, is identified and characterized. The dysferlin gene is normally expressed in skeletal muscle cells and is selectively mutated in several families with the hereditary muscular dystrophies, e.g., Miyoshi myopathy  
15 (MM) and limb girdle muscular dystrophy-2B (LGMD2B). These characteristics of dysferlin render it a candidate disease gene for both MM and LGMD2B. An additional novel protein, brain-specific dysferlin, has also been identified. Defects in brain-specific dysferlin may  
20 predispose to selected disorders of the central nervous system. Moreover, the expression of brain-specific dysferlin may be important as a marker for normal neural development (e.g., *in vivo* or in neural cells in culture). Manipulation of levels of expression of brain-  
25 specific dysferlin, and of the type of expressed brain-specific dysferlin is of use for analyzing the function of brain-specific dysferlin and related dysferlin-associated molecules.

The invention features an isolated DNA which  
30 includes a nucleotide sequence hybridizing under stringent hybridization conditions to a strand of SEQ ID NO:3 or SEQ ID NO:117.

- 3 -

The invention also features an isolated DNA including a nucleotide sequence selected from SEQ ID NOS:4-12.

Also within the invention is an isolated DNA  
5 comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:22-30.

Also within the invention is a single stranded oligonucleotide of 14-50 nucleotides in length having a nucleotide sequence identical to a portion of a strand of  
10 SEQ ID NO:3.

Also within the invention is a pair of PCR primers consisting of:

(a) a first single stranded oligonucleotide consisting of 14-50 contiguous nucleotides of the sense  
15 strand of SEQ ID NO:117; and

(b) a second single stranded oligonucleotide consisting of 14-50 contiguous nucleotides of the antisense strand of SEQ ID NO:117, wherein the sequence of at least one of the oligonucleotides is identical to a  
20 portion of a strand of SEQ ID NO:3, and the first oligonucleotide is not complementary to the second oligonucleotide.

Also within the invention is a pair of single stranded oligonucleotides selected from of SEQ ID NOS  
25 130-231, SEQ ID NO:110, and SEQ ID NO:112.

Also within the invention is an isolated DNA including a nucleotide sequence that encodes a protein that shares at least 70% sequence identity with SEQ ID NO:2, or a complement of the nucleotide sequence.

30 Also within the invention is an isolated DNA including a nucleotide sequence which hybridizes under stringent hybridization conditions to a strand of a nucleic acid, the nucleic acid having a sequence selected from SEQ ID NOS:31-79 and 90-101.

- 4 -

Also within the invention is a single stranded oligonucleotide of 14-50 nucleotides in length having a nucleotide sequence which is identical to a portion of a strand of a nucleic acid selected from SEQ ID NOS:31-79 and 90-100.

Also within the invention is a pair of PCR primers consisting of:

(a) a first single stranded oligonucleotide consisting of 14-50 contiguous nucleotides of the sense strand of a nucleic acid selected from SEQ ID NOS:31-85; and

(b) a second single stranded oligonucleotide consisting of 14-50 contiguous nucleotides of the antisense strand of a nucleic acid selected from SEQ ID NOS:31-85, wherein the sequence of at least one of the oligonucleotides includes a sequence identical to a portion of a strand of a nucleic acid selected from SEQ ID NOS: 31-79 and 90-100, and the first oligonucleotide is not complementary to the second oligonucleotide.

Also within the invention is a pair of single stranded oligonucleotides selected from SEQ ID NOS 101-116, SEQ ID NOS 184-185, SEQ ID NOS 188-191, SEQ ID NOS 210-213, and SEQ ID NOS 216-217.

Also within the invention is a substantially pure protein that has an amino acid sequence sharing at least 70% sequence identity with SEQ ID NO:2.

Also within the invention is a substantially pure protein the sequence of which includes amino acid residues 1-500, 501-1000, 1001-1500, or 1501-2080 of SEQ ID NO:2.

Also within the invention is a substantially pure protein including the amino acid sequence of SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88, or SEQ ID NO:89.

In another aspect, the invention features a transgenic non-human mammal having a transgene disrupting

- 5 -

or interfering with the expression of a dysferlin gene, the transgene being chromosomally integrated into the germ cells of the animal.

Another embodiment of the invention features a method of decreasing the symptoms of muscular dystrophy in a mammal by introducing into a cell of the mammal (e.g., a muscle cell or a muscle precursor cell) an isolated DNA which hybridizes under stringent hybridization conditions to a strand of SEQ ID NO:3.

10 Another aspect of the invention provides a method for identifying a patient, a fetus, or a pre-embryo at risk for having a dysferlin-related disorder by (a) providing a sample of genomic DNA from the patient, fetus, or pre-embryo; and (b) determining whether the  
15 sample contains a mutation in a dysferlin gene.

In another aspect, the invention provides a method for identifying a patient, a fetus, or a pre-embryo at risk for having a dysferlin-related disorder by (a) providing a sample including dysferlin mRNA from the  
20 patient, fetus, or pre-embryo; and (b) determining whether the dysferlin mRNA contains a mutation.

Methods of identifying mutations in a dysferlin sequence are useful for predicting (e.g., predicting whether an individual is at risk for developing a  
25 dysferlin-related disorder) or diagnosing disorders associated with dysferlin, e.g., MM and LGMD2B. Such methods can also be used to determine if an individual, fetus, or a pre-embryo is a carrier of a dysferlin mutation, for example in screening procedures. Methods  
30 which distinguish between different dysferlin alleles (e.g., a mutant dysferlin allele and a normal dysferlin allele) can be used to determine carrier status.

The invention also features an isolated nucleic acid comprising a nucleotide sequence which hybridizes  
35 under stringent hybridization conditions to nucleic acids

- 6 -

3284-3720 of SEQ ID NO:232, or the complement of the nucleotide sequence. An isolated nucleic acid including a nucleotide sequence identical to the sequence of nucleotides 3284-3720 of SEQ ID NO:232, or a complement  
5 of the nucleotide sequence is also a feature of the invention. The isolated nucleic acid can include the entire sequence of SEQ ID NO:232 or the complement of SEQ ID NO:232.

Another aspect of the invention features an  
10 isolated polypeptide that includes: a) at least 15 contiguous amino acids of the polypeptide comprising amino acids 1-24 of SEQ ID NO:233, b) a naturally occurring allelic variant of a polypeptide comprising amino acids 1-24 of SEQ ID NO:233, or c) an amino  
15 acid sequence which is encoded by a nucleic acid molecule which hybridizes under stringent conditions to nucleotides 3284-3720 of SEQ ID NO:232. The polypeptide of this aspect can include the entire sequence of SEQ ID NO:233.

20 Also included in the invention is a vector comprising the nucleic acid of claim 44 and a cell that contains the vector. Another aspect of the invention features a method of making a polypeptide by culturing the cell which contains the vector.

25 The invention also features an antibody which specifically binds to a polypeptide of such as those described above. The antibody can bind to a polypeptide selected from amino acids 253-403 of SEQ ID NO:233, amino acids 624-865 of SEQ ID NO:233, and amino acids 1664-1786  
30 of SEQ ID NO:233. Antibodies of the invention can be monoclonal or polyclonal antibodies.

An "isolated DNA" is DNA which has a naturally occurring sequence corresponding to part or all of a given gene but is free of the two genes that normally  
35 flank the given gene in the genome of the organism in

- 7 -

which the given gene naturally occurs. The term therefore includes a recombinant DNA incorporated into a vector, into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote. It also includes a separate molecule such as a cDNA, a genomic fragment, a fragment produced by polymerase chain reaction (PCR), or a restriction fragment, as well as a recombinant nucleotide sequence that is part of a hybrid gene, i.e., a gene encoding a fusion protein. The term excludes intact chromosomes and large genomic segments containing multiple genes contained in vectors or constructs such as cosmids, yeast artificial chromosomes (YACs), and P1-derived artificial chromosome (PAC) contigs.

15 A "noncoding sequence" is a sequence which corresponds to part or all of an intron of a gene, or to a sequence which is 5' or 3' to a coding sequence and so is not normally translated.

An expression control sequence is "operably linked" to a coding sequence when it is within the same nucleic acid and can control expression of the coding sequence.

A "protein" or "polypeptide" is any chain of amino acids linked by peptide bonds, regardless of length or post-translational modification, e.g., glycosylation or phosphorylation.

As used herein, the term "percent sequence identity" means the percentage of identical subunits at corresponding positions in two sequences when the two sequences are aligned to maximize subunit matching, i.e., taking into account gaps and insertions. For purposes of the present invention, percent sequence identity between two polypeptides is to be determined using the Gap program and the default parameters as specified therein.

35 The Gap program is part of the Sequence Analysis Software

- 8 -

Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705.

The algorithm of Myers and Miller, CABIOS (1989) can also be used to determine whether two sequences are similar or identical. Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used.

As used herein, the term "stringent hybridization conditions" means the following DNA hybridization and wash conditions: hybridization at 60°C in the presence of 6 x SSC, 0.5% SDS, 5 x Denhardt's Reagent, and 100 µg/ml denatured salmon sperm DNA; followed by a first wash at room temperature for 20 minutes in 0.5 x SSC and 0.1% SDS and a second wash at 55°C for 30 minutes in 0.2 x SSC and 0.1% SDS.

A "substantially pure protein" is a protein separated from components that naturally accompany it. The protein is considered to be substantially pure when it is at least 60%, by dry weight, free from the proteins and other naturally-occurring organic molecules with which it is naturally associated. Preferably, the purity of the preparation is at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight. A substantially pure dysferlin protein can be obtained, for example, by extraction from a natural source, by expression of a recombinant nucleic acid encoding a dysferlin polypeptide, or by chemical synthesis. Purity can be measured by any appropriate method, e.g., column chromatography, polyacrylamide gel electrophoresis, or HPLC analysis. A chemically synthesized protein or a recombinant protein produced in a cell type other than



- 9 -

the cell type in which it naturally occurs is, by definition, substantially free from components that naturally accompany it. Accordingly, substantially pure proteins include those having sequences derived from eukaryotic organisms but which have been recombinantly produced in *E. coli* or other prokaryotes.

An antibody that "specifically binds" to an antigen is an antibody that recognizes and binds to the antigen, e.g., a dysferlin polypeptide, but which does not substantially recognize and bind to other molecules in a sample (e.g., a biological sample) which naturally includes the antigen, e.g., a dysferlin polypeptide. An antibody that "specifically binds" to dysferlin is sufficient to detect a dysferlin polypeptide in a biological sample using one or more standard immunological techniques (for example, Western blotting or immunoprecipitation).

A "transgene" is any piece of DNA, other than an intact chromosome, which is inserted by artifice into a cell, and becomes part of the genome of the organism which develops from that cell. Such a transgene may include a gene which is partly or entirely heterologous (i.e., foreign) to the host organism, or may represent a gene homologous to an endogenous gene of the organism.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention. The present materials, methods, and examples are illustrative only and not intended to be limiting. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present

- 10 -

specification, including definitions, will control. All the sequences disclosed in the sequence listing are meant to be double-stranded except the sequences of oligonucleotides.

- 5 Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

#### Brief Description of the Drawings

Fig. 1A is a physical map of the MM locus. Arrows  
10 indicate the five new polymorphic markers and filled, vertical rectangular boxes indicate the previously known polymorphic markers. The five ESTs that are expressed in skeletal muscle are highlighted in bold. Detailed information on the minimal tiling path of the PAC contig  
15 spanning the MM/LGMD2B region is provided in Liu et al., 1998, *Genomics* 49:23-29. The minimal candidate MM region is designated by the solid bracket (top) and compared to the previous candidate region (dashed bracket). TGFA and ADD2 are transforming growth factor alpha and  $\beta$ -adducin  
20 2.

Fig. 1B is a representation of the dysferlin cDNA clones. The probes used in the three successive screens are shown in bold (130347, cDNA10, A27-F2R2). The two most 5' cDNA clones are also shown (B22, B33). The 6.9  
25 kb cDNA for dysferlin (SEQ ID NO:1) is illustrated at the bottom with start and stop codons as shown.

Fig. 1C is a representation of the predicted dysferlin protein. The locations of four C2 domains (SEQ ID NOs: 86-89) are indicated by stippled boxes,  
30 while the putative transmembrane region is hatched. Vertical lines above the cDNA denote the positions of the mutations in Table 2; the associated labels indicate the phenotypes (MM - Miyoshi myopathy; LGMD - limb girdle

- 11 -

muscular dystrophy; DMAT - distal myopathy with anterior tibial onset).

Fig. 2 is the sequence of the predicted 2,080 amino acids of dysferlin (SEQ ID NO:2). The predicted  
5 membrane spanning residues are in bold at the carboxy terminus (residues 2047-2063). Partial C2 domains are underlined. Bold, underlined sequences are putative nuclear targeting residues. Possible membrane retention sequences are enclosed within a box.

10 Fig. 3 is a comparison of the Kyle-Doolittle hydrophobicity plots of the dysferlin protein and fer-1. On the Y-axis, increasing positivity corresponds to increasing hydrophobicity. Both proteins have a single, highly hydrophobic stretch at the carboxy terminal end  
15 (arrow). Both share regions of relative hydrophilicity approximately at residue 1,000 (arrowhead).

Fig. 4 is a SSCP analysis of a representative pedigree with dysferlin mutations. Each member of the pedigree is illustrated above the corresponding SSCP  
20 analysis. For each affected individual (solid symbols) shifts are evident in alleles 1 and 2, corresponding respectively to exons 36 and 54. As indicated, the allele 1 and 2 variants are transmitted respectively from the mother and the father. The two affected daughters in  
25 this pedigree have the limb girdle muscular dystrophy (LGMD) phenotype while their affected brother has a pattern of weakness suggestive of Miyoshi myopathy (MM).

Fig. 5 is a representation of the genomic structure of dysferlin. The 55 exons of the dysferlin  
30 gene and their corresponding SEQ ID NOs are indicated below the 6911 bp cDNA (solid line). The cDNA sequences corresponding to SEQ ID NO:1 and SEQ ID NO:3 are shown relative to the 6911 bp cDNA.

Figs. 6A-B are the cDNA sequence of brain-specific  
35 dysferlin (SEQ ID NO:232) and the predicted amino acid

- 12 -

sequence (in single-letter code) of brain-specific  
dysferlin (SEQ ID NO:233).

#### Detailed Description

The Miyoshi myopathy (MM) locus maps to human  
5 chromosome 2p12-14 between the genetic markers D2S292 and  
D2S286 (Bejaoui et al., 1995, *Neurology* 45:768-72).  
Further refined genetic mapping in MM families placed the  
MM locus between markers GGAA-P7430 and D2S2109 (Bejaoui  
et al., 1998, *Neurogenetics* 1:189-96). Independent  
10 investigation has localized the limb-girdle muscular  
dystrophy (LGMD-2B) to the same genetic interval (Bashir  
et al., 1994, *Hum. Molec. Genetics* 3:455-57; Bashir et  
al., 1996, *Genomics* 33:46-52; Passos-Bueno et al., 1995,  
*Genomics* 27:192-95). Furthermore, two large, inbred  
15 kindreds have been described whose members include both  
MM and LGMD2B patients (Weiler et al., 1996, *Am. J. Hum.*  
*Genet.* 59:872-78; Illarioshkin et al., 1997, *Genomics*  
42:345-48). In these familial studies, the disease  
gene(s) for both MM and LGMD2B mapped to essentially the  
20 same genetic interval. Moreover, in both pedigrees,  
individuals with MM or LGMD2B phenotypes share the same  
haplotypes. This raises the intriguing possibility that  
the two diseases may arise from the same gene defect and  
that a particular disease phenotype is the result of  
25 modification by additional factors.

A 3-Mb PAC contig spanning the entire MM/LGMD2B  
candidate region was recently constructed to facilitate  
the cloning of the MM/LGMD2B gene(s) (Liu et al., 1998,  
*Genomics* 49:23-29). This high resolution PAC contig  
30 resolved the discrepancies of the order of markers in  
previous studies (Bejaoui et al., 1998, *Neurogenetics*  
1:189-96; Bashir et al., 1996, *Genomics* 33:46-52; Hudson  
et al., 1995, *Science* 270:1945-54). The physical size of  
the PAC contig also indicated that the previous minimal

- 13 -

size estimation based on YAC mapping data was significantly underestimated.

#### Identification of Repeat Sequences and Repeat Typing

The PAC contig spanning the MM/LGMD2B region (Liu et al., 1998, *Genomics* 49:23-29) was used as a source for the isolation of new informative markers to narrow the genetic interval of the disease gene(s). DNA from the PAC clones spanning the MM/LGMD2B region was spotted onto Hybond N+™ membrane filters (Amersham, Arlington Heights, IL). The filters were hybridized independently with the following  $\gamma$ -<sup>32</sup>P (Du Pont, Wilmington, DE) labeled repeat sequences: (1) (CA)<sub>15</sub>; (2) pool of (ATT)<sub>10</sub>, (GATA)<sub>8</sub> and (GGAA)<sub>8</sub>; (3) pool of (GAAT)<sub>8</sub>, (GGAT)<sub>8</sub> and (GTAT)<sub>8</sub>; and (4) pool of (AAG)<sub>10</sub> and (ATC)<sub>10</sub>. Hybridization and washing of the filters were carried out at 55°C following standard protocols (Sambrook et al., 1989, *Molecular Cloning: A Laboratory Manual* (2nd Edition), Cold Spring Harbor Press, N.Y.).

Miniprep DNAs of PAC clones containing repeat sequences were digested with restriction enzymes *Hind*III and *Pst*I and ligated into pBluescript II (KS+) vector which is (Stratagene, La Jolla, CA) digested with the same enzymes. Filters of the PAC subclones were hybridized to the  $\gamma$ -<sup>32</sup>P labeled repeats that detected the respective PACs. For clones with an insert size greater than 1 kb the repeat sequences of which could not be identified by a single round of sequencing, the inserts were further subcloned by digestion with *Hae*III and ligation in *Eco*RV-digested pZero-2.1 vector (Invitrogen, Inc., Carlsbad, CA). Miniprep DNAs of the positive subclones were subjected to manual dideoxy sequencing with Sequenase™ enzyme (US Biochemicals, Inc., Cleveland, OH). Primer pairs for amplifying the repeat sequences were selected using the computer program Oligo (Version

- 14 -

4.0, National Biosciences, Inc., Plymouth, MN). Primer sequences are shown in Table 1.

TABLE 1

## New Polymorphic Markers Mapped to the MM/LGMD2B Region

| <u>Marker</u>          | <u>Repeat</u> | <u>Primers (5' to 3')</u>  | <u>Annealing<br/>T<sub>m</sub> (°C)</u> | <u>Size in<br/>PAC (bp)</u> | <u>No. of<br/>alleles<sup>1</sup></u> | <u>Het<sup>2</sup></u> |
|------------------------|---------------|--|---|-----------------------------|---------------------------------------|------------------------|
| PAC3-H52               | CA            | GATCTAACCCCTGCTCACC<br>(SEQ ID NO:120)<br>CTGGTGTGTTGCAGAGCGCTG<br>(SEQ ID NO:121)   | 57                                      | 138                         | 10                                    | 0.82                   |
| Cy172-H32 <sup>3</sup> | CCAT          | CCTCTCTTCTGCTGCTTCAG<br>(SEQ ID NO:122)<br>TGTGCTGTGTTCCACCTTCGT<br>(SEQ ID NO:123)  | 56                                      | 199                         | 7                                     | 0.72                   |
| PAC35-PH2              | CAT           | TCCAAATAGAAATGCCTGAAC<br>(SEQ ID NO:124)<br>AGGTATCACCTCCAAGTGTG<br>(SEQ ID NO:125)  | 56                                      | 161                         | 5                                     | 0.30                   |
| PAC16-H41              | Complex       | TACCAGCTTCAGAGCTCCCTG<br>(SEQ ID NO:126)<br>TTGATCAGGGTGCTCTTGG<br>(SEQ ID NO:127)   | 58                                      | 280                         | 4                                     | 0.41                   |
| Cy7-PH3                | AAGG          | GGAGAATTGCTTGAACCCAG<br>(SEQ ID NO:128)<br>TGGCTAATGATGTTGAACATTT<br>(SEQ ID NO:129) | 56                                      | 211                         | 4                                     | 0.32                   |

<sup>1</sup> Observed in 50 unrelated caucasians.<sup>2</sup> Heterozygosity index.<sup>3</sup> Located within intron 2 of the *dysferlin* gene.

All oligonucleotides were synthesized by Integrated DNA Technologies, Inc. (Coralville, IA). PCR typing of the repeat markers followed previously described protocols (Bejaoui et al., 1995, Neurology 45:768-772).

- 16 -

Identification of Repeat Markers and Haplotype Analysis

After hybridization with labeled repeat oligos, 17 different groups of overlapping PACs were identified that contained repeat sequences. Some groups contained previously identified repeat markers. For example, five groups of PACs were positively identified by a pool of repeat probes including (ATT)<sub>10</sub>, (GATA)<sub>8</sub>, and (GGAA)<sub>8</sub>. Of these, three groups contained known markers GGAA-P7430 (GGAA repeat), D2S1394 (GATA repeat) and D2S1398 (GGAA repeat) (Hudson et al., 1992, *Nature* 13:622-29; Gastier et al., 1995, *Hum. Molecular Genetics* 4:1829-36). No attempt was made to isolate new repeat markers from these PACs and they were not further analyzed. Similarly, seven groups of PACs that contained known CA repeat markers were excluded. Seven groups of PACs that contained unidentified repeats were retained for further analysis. For each group, the PAC containing the smallest insert was selected for subcloning. Subclones were re-screened and positive clones were sequenced to identify repeats. In total, seven new repeat sequences were identified within the MM/LGMD2B PAC contig. Of these, five are polymorphic within the population that was tested. The information for these five markers is summarized in Table 1. Based on the PAC contig constructed previously across the MM candidate locus (Liu et al., 1998, *Genomics* 48:23-29), the five new markers and ten previously published polymorphic markers were placed in an unambiguous order (Fig. 1).

These markers were analyzed in a large, consanguineous MM family (Bejaoui et al., 1995, *Neurology* 45: 768-72; Bejaoui et al., 1998, *Neurogenetics* 1:189-96). Because MM is a recessive condition, the locus can be defined by identifying regions of the genome that show homozygosity in affected individuals. Conversely, because of the high penetrance of this adult-onset



- 17 -

condition, unaffected adult individuals are not expected to be homozygous by descent across the region. Analysis of haplotype homozygosity in this pedigree indicates that the disease gene lies between markers D2S2111 and PAC3-H52. Based on the PAC mapping data, the physical distance for this interval is approximately 2.0 Mb. No recombination events were detected between four informative markers (markers cy172-H32 to PAC16-H41) and the disease locus in family MM-21 (Fig. 1A).

#### 10 Identification of Five Muscle-Expressed ESTs

Twenty-two ESTs and two genes (transforming growth factor alpha [TGF $\alpha$ ] and beta-adducin [ADD2]) were previously mapped to the MM/LGMD2B PAC contig (Fig. 1A) (Liu et al., 1998, *Genomics* 48:23-29). Two  $\mu$ l (approximately 0.1 ng/ $\mu$ l) of Marathon-ready™ skeletal muscle cDNA (Clontech, Palo Alto, CA) were used as template in a 10  $\mu$ l PCR reaction for analysis of muscle expression of ESTs. The PCR conditions were the same as for the PCR typing of repeat markers. PCR analysis of skeletal muscle cDNA indicated that five of these ESTs (A006G04, stSG1553R, WI-14958, TIGR-A004Z44 and WI-14051) map within the minimal genetic MM interval of MM and are expressed in skeletal muscle.

Probes were selected corresponding to each of these five ESTs for Northern blot analysis. cDNA clones (130347, 48106, 172575, 184080, and 510138) corresponding to the five ESTs that are expressed in muscle (respectively TIGR-A004Z44, WI-14051, WI-14958, stSG1553R and A006G04) were selected from the UniGene database (<http://www.ncbi.nlm.nih.gov/UniGene/>) and obtained from Genome Systems, Inc. (St. Louis, MO). The cDNA probes were first used to screen the MM/LGMD2B PAC filters to confirm that they mapped to the expected position in the MM/LGMD2B contig.

- 18 -

A Northern blot (Clontech) of multiple human tissues was sequentially hybridized to the five cDNA probes and a control  $\beta$ -actin cDNA at 65°C following standard hybridization and washing protocols (Sambrook et al., *supra*). Between hybridizations, probes were removed by boiling the blot at 95-100°C for 4-10 min with 0.5% SDS. The blot was then re-exposed for 24 h to confirm the absence of previous hybridization signals before proceeding with the next round of hybridization.

10 The tissue distribution, intensity of the signals and size of transcripts detected by the five cDNA probes varied. Probes corresponding to ESTs stSG1553R, TIGR-A004Z44 and WI-14958 detected strong signals in skeletal muscle. In addition, the cDNA corresponding to TIGR-  
15 A004Z44 detected a 3.6-3.8 kb brain-specific transcript instead of the 8.5 kb message that was present in other tissues. It is likely that these five ESTs correspond to different genes since the corresponding cDNA probes used for Northern analysis derive from the 3' end of messages,  
20 map to different positions in the MM/LGMD2B contig (Fig. 1A), and differ in their expression patterns.

Current database analysis suggests that three of these ESTs (stSG1553R, WI-14958 and WI-14051) do not match any known proteins (Schuler et al., 1996, Science  
25 274:540-46). A006G04 has weak homology with a protein sequence of unknown function that derives from *C. elegans*. TIGR-A004Z44 has homology only to subdomains present within protein kinase C. Because the five genes corresponding to the ESTs are expressed in skeletal  
30 muscle and map within the minimal genetic interval of the MM/LGMD2B gene(s), they are candidate MM/LGMD2B gene(s).

- 19 -

Cloning of Dysferlin cDNA

EST TIGR-A004Z44 gave a particularly strong skeletal muscle signal on the Northern blot. Moreover, it is bracketed by genetic markers that show no recombination with the disease phenotype in family MM-21 (Fig. 1). The corresponding transcript was therefore cloned and analyzed as a candidate MM gene. From the Unigene database, a cDNA IMAGE clone (130347, 979 bp) was identified that contained the 483 bp EST TIGR-A004Z44.

10 Approximately  $1 \times 10^6$  recombinant clones of a  $\lambda$ gt11 human skeletal muscle cDNA library (Clontech) were plated and screened following standard techniques (Sambrook et al., *supra*). The initial library screening was performed using the insert released from the clone 130347 that  
15 contains EST TIGR-A0044Z44, corresponding to the 3' end of the gene. Positive phages were plaque purified and phage DNA was isolated according to standard procedures (Sambrook et al., *supra*). The inserts of the positive clones were released by *EcoRI* digestion of phage DNA and  
20 subsequently subcloned into the *EcoRI* site of pBluescript II (KS+) vector (Stratagene).

Fifty cDNA clones were identified when a human skeletal muscle cDNA library was screened with the 130347 cDNA. Clone cDNA10 with the largest insert (~6.5 kb)  
25 (Fig. 1B) was digested independently with *BamHI* and *PstI* and further subcloned into pBluescript vector. Miniprep DNA of cDNA clones and subclones of cDNA10 was prepared using the Qiagen plasmid Miniprep kit (Valencia, CA). Sequencing was carried out from both ends of each clone  
30 using the SequiTherm EXCEL™ long-read DNA sequencing kit (Epicenter, Madison, WI), fluorescent-labeled M13 forward and reverse primers, and a LI-COR sequencer (Lincoln, NE). Assembly of cDNA contigs and sequence analysis were performed using Sequencer software (Gene Codes  
35 Corporation, Inc., Ann Arbor, MI).

- 20 -

Two additional screens, first with the insert of cDNA10 and then a 683 bp PCR product (A27-F2R2) amplified from the 5' end of the cDNA contig, identified 87 additional cDNA clones. Clones B22 and B33 extended the 5' end by 94 and 20 bp, respectively. The compiled sequence allowed for the generation of a sequence of 6.9 kb (SEQ ID NO:1) (with 10-fold average coverage).

Although the 5' end of the gene has not been further extended to the 8.5 kb predicted by Northern analysis, an open reading frame (ORF) of 6,243 bp has been identified within this 6.9 kb sequence. This ORF is preceded by an in-frame stop codon and begins with the sequence cgcaagcATGCTG (SEQ ID NO:118); five of the first seven bp are consistent with the Kozak consensus sequence for a start codon (Kozak, 1989, *Nucl. Acids Res.* 15:8125-33; Kozak, 1989, *J. Cell. Biol.* 108:229-41). An alternate start codon, in the same frame, +75 bp downstream, appears less likely as a start site GAGACGATGGGG (SEQ ID NO:119). Thus, the entire coding region of this candidate gene is believed to have been identified, as represented by the 6.9 kb sequence contig.

#### Isolation of the Brain-Specific Dysferlin Isoform

##### Identification of the brain-specific isoform of dysferlin

A brain-specific isoform of dysferlin was identified using Northern blot analysis of poly(A+)RNA derived from multiple human adult tissues probed with radiolabeled full-length dysferlin cDNA subclones. A prominent 7.2 kb transcript was detected on Northern blots in skeletal muscle, heart, placenta, lung, and kidney, while a distinct but equally prominent 3.6 kb-3.8 kb transcript was identified exclusively in the brain. Using long exposures, a faint 7.2 kb mRNA was also detected in the

- 21 -

brain. This finding suggested that the shorter brain isoform was likely to be a tissue-specific splice variant of the dysferlin gene. To test this hypothesis, a human brain cDNA library (Stratagene) was screened for the  
5 dysferlin brain isoform.

Cloning of the brain-specific dysferlin isoform

To identify probes that hybridize to the brain-specific dysferlin sequence and so could be used for library screening, fragments of the full-length dysferlin  
10 cDNA clone (derived from a skeletal muscle cDNA library) were generated using restriction enzymes. The fragments were about 1 kb in length and were analyzed by hybridization to a Northern blot that included brain RNA. Sequences suitable for library screening were those that  
15 hybridized to the 3.6-3.8 kb brain-specific transcript. A region of the 3' end of the dysferlin cDNA sequence that is approximately 3 kb in length was identified as hybridizing to brain mRNA. DNA containing sequence from this region was used as a probe for hybridization  
20 screening of a human brain cDNA library (Stratagene).

The human brain cDNA library was plated out and screened using standard procedures. Of the approximately 720,000 plaques screened, 63 primary positive clones were identified. Of these, 20 clones were selected for  
25 further analysis involving standard methods of hybridization, restriction enzyme mapping, and sequencing. The primary positive clones shared regions of overlap with each other.

Sequencing of positive clones, provided 3671  
30 nucleotides of the brain-specific dysferlin sequence (SEQ ID NO:232; Figure 6A-B). The identified sequence corresponds closely to the size of the brain-specific dysferlin transcript detected on Northern blots. With the exception of the 5' region of the sequence, the

- 22 -

brain-specific sequence is identical to about 3.1 kb of the dysferlin sequence (from nucleotide 3722 to 6904 of the dysferlin sequence). In the dysferlin gene, position 3722 corresponds to the start of exon 32. This finding  
5 is consistent with the hypothesis that the brain isoform is a splice-variant of the dysferlin gene. At the 5' end of the brain isoform, 489 nucleotides are unique to brain-specific dysferlin. The amino acid sequence encoded by the brain dysferlin nucleic acid sequence (SEQ  
10 ID NO:233; Figure 6) contains a unique sequence with an initiation codon within a Kozak consensus sequence. The nucleic acid sequence unique to brain-specific dysferlin encodes a novel 24 amino acid sequence.

#### Identification of Mutations in Miyoshi Myopathy

15 Two strategies were used to determine whether this 6.9 kb cDNA (SEQ ID NO:1) is mutated in MM. First, the genomic organization of the corresponding gene was determined and the adjoining intronic sequence at each of the 55 exons which make up the cDNA was identified. To  
20 identify exon-intron boundaries within the gene, PAC DNA was extracted with the standard Qiagen -Mini Prep protocol. Direct sequencing was performed with DNA Sequence System (Promega, Madison, WI) using <sup>32</sup>P end-labeled primers (Benes et al., 1997, *Biotechniques* 23:98-  
25 100). Exon-intron boundaries were identified as the sites where genomic and cDNA sequences diverged. Second, in patients for whom muscle biopsies were available, RT-PCR was also used to prepare cDNA for the candidate gene from the muscle biopsy specimen.

30 Single strand conformational polymorphism analysis (SSCP) was used to screen each exon in patients from 12 MM families. Putative mutations identified in this way were confirmed by direct sequencing from genomic DNA using exon-specific intronic primers. Approximately 20

- 23 -

ng of total genomic DNA from immortalized lymphocyte cell lines were used as a template for PCR amplification analysis of each exon using primers (below) located in the adjacent introns. SSCP analysis was performed as previously described (Aoki et al., 1998, *Ann. Neurol.* 43:645-53). In patients for whom muscle biopsies were available, mRNA was isolated using RNA-STAT-60™ (Tel-Test, Friendswood, TX) and first-strand cDNA was synthesized from 1-2 µg total RNA with MMLV reverse transcriptase and random hexamer primers (Life Technologies, Gaithersburg, MD). Three µl of this product were used for PCR amplification. Eight sets of primers were designed for muscle cDNA, and overlapping cDNA fragments suitable for SSCP analysis were amplified. After initial denaturation at 94°C for 2 min, amplification was performed using 30 cycles at 94°C for 30 s, 56°C for 30 s, and 72°C for 60 s. The sequences of polymorphisms detected by SSCP analysis were determined by the dideoxy termination method using the Sequenase kit (US Biochemicals). In some instances, the base pair changes predicted corresponding changes in restriction enzyme recognition sites. Such alterations in restriction sites were verified by digesting the relevant PCR products with the appropriate restriction enzymes.

Primer pairs used for SSCP screening and exon sequencing are as follows:

- (1) exon 3, F3261 5'-tctcttctcctagaggccatag-3' (SEQ ID NO: 101) and R326 5'-ctgttcctccccatcgtctcatgg-3' (SEQ ID NO: 102);
- (2) exon 20, F3121 5'-gctcctcccgtgaccctctg-3' (SEQ ID NO: 103) and R3121 5'-gggtcccagccaggagcactg-3' (SEQ ID NO: 104);
- (3) exon 36, F2102 5'-cccctctcaccatctcctgatgtg-3' (SEQ ID NO: 105) and R2111 5'-tggcttcaccttcctctacctcgg-3' (SEQ ID NO: 106);

- 24 -

- (4) exon 49, F1081 5'-tcctttggtaggaaatctaggtgg-3'  
(SEQ ID NO: 107) and R1081 5'-ggaagctggacaggcaagagg-3'  
(SEQ ID NO: 108);
- (5) exon 50, F1091 5'-atatactgtgttggaatcttaatgag-3'  
5 (SEQ ID NO: 109) and R1091 5'-gctggcaccacaggggaatcgg-3'  
(SEQ ID NO: 110);
- (6) exon 51, F1101 5'-ctttgcttccttgcataccttctctg-3'  
(SEQ ID NO: 111) and R1101 5'-agcccccatgtgcagaatggg-3'  
(SEQ ID NO: 112);
- 10 (7) exon 52, F1111 5'-ggcagtgatcgagaaacccgg-3' (SEQ  
ID NO: 113) and R1111 5'-catgccctccactggggctgg-3' (SEQ ID  
NO: 114);
- (8) exon 54, F1141 5'-ggatgcccagttgactccggg-3' (SEQ ID  
NO: 115) and R1141 5'-ccccaccacagtgtcgtcagg-3' (SEQ ID NO:  
15 116);
- (9) exon 29, F3031 5'-aagtgccaaagcaatgagtgaccgg-3' (SEQ  
ID NO: 184) and R3021 5'-ctcactcccacccaccacctg-3' (SEQ ID  
NO: 185);
- (10) exon 31, F2141 5'-gaatctgccataaccagcttcgtg-3' (SEQ  
20 ID NO: 188) and R2141 5'-tatcaccccatagaggcctcgaag-3' (SEQ ID  
NO: 189);
- (11) exon 32, F2981 5'-cagccactcactctggcacctctg-3' (SEQ  
ID NO: 190) and R2981 5'-agcccacagtctctgactctcctg-3' (SEQ ID  
NO: 191);
- 25 (12) exon 43, F2031 5'-cagccaaaccatatcaacaatg-3' (SEQ  
ID NO: 210) and R2021 5'-ctggggaggtgagggctctag-3' (SEQ ID  
NO: 211);
- (13) exon 44, F2011 5'-gaagtgttttgtctcctcctc-3' (SEQ ID  
NO: 212) and R2011 5'-gcaggcagccagccccatc-3' (SEQ ID NO:  
30 213);
- (14) exon 46, F1041 5'-ctcgtctatgtcttgtgcttgctc-3' (SEQ  
ID NO: 216) and R1051 5'-caccatgggttggggcatgtgg-3' (SEQ ID  
NO: 217).



- 25 -

These primers were used in SSCP screening and exon sequencing, and identified eighteen different mutations in fifteen families (Table 2).

Table 2  
Mutations in Dysferlin in Distal Myopathy and LGMD<sup>1</sup>

| Name      | Nucleotide<br>Change                  | Exon | Consequence          | Origin  | Family<br>name | Allele | Change of<br>restricti<br>on site |
|-----------|---------------------------------------|------|----------------------|---------|----------------|--------|-----------------------------------|
| Mutations |                                       |      |                      |         |                |        |                                   |
| 5 537insA | ins of A<br>at 537                    | 3    | Frameshift           | Arabic  | MM59           | Hom    | no change                         |
| Q605X     | <u>C</u> AG to <u>T</u> AG<br>at 2186 | 20   | Stop at 605          | French  | MM67           | Hom    | -Pst I,<br>-Fnu 4H I <sup>1</sup> |
| I1298V    | <u>A</u> TC to <u>G</u> TC<br>at 4265 | 36   | Amino acid<br>change | Italian | MM,<br>LGMD56  | Het    | -BamHI,<br>-BstYI;<br>+Ava II     |
| E1883X    | <u>G</u> AG to <u>T</u> AG<br>at 5870 | 49   | Stop at<br>1883      | English | MM8            | Het    | no change                         |
| H1857R    | <u>C</u> AT to <u>C</u> GT<br>at 5943 | 50   | Amino acid<br>change | English | MM50           | Het    | no change                         |

|                    |                                       |    |                      |          |        |     |                                       |
|--------------------|---------------------------------------|----|----------------------|----------|--------|-----|---------------------------------------|
| 5966delG           | del of G<br>at 5966                   | 50 | Frameshift           | Spanish  | DMAT71 | Hom | no change                             |
| 5966delG           | del of G<br>at 5966                   | 50 | Frameshift           | Spanish  | MM75   | Hom | no change                             |
| 6071/6072de<br>1AG | del of AG<br>at<br>6071/6072          | 51 | Frameshift           | English  | MM58   | Het | no change                             |
| 5 6319+1G to<br>A  | Ggt to Gat<br>at 6319+1               | 52 | 5' splice<br>site    | English  | MM8    | Het | no change                             |
| R2042C             | <u>C</u> GT to <u>T</u> GT<br>at 6497 | 54 | Amino acid<br>change | Italian  | MM56   | Het | -Fnu4HI                               |
| R1046H             | <u>C</u> GC to <u>C</u> AG<br>at 3510 | 29 | Amino acid<br>change | Japanese | MM10   | Hom | -HinPI,<br>-Fsp I                     |
| 3746delG           | del of G<br>at 3746                   | 31 | Frameshift           | Japanese | MM17   | Hom | -MboII                                |
| 10 Q1160X          | <u>C</u> AG to <u>T</u> AG<br>at 3851 | 32 | Stop at<br>1160      | Mexican  | MM46   | Hom | -ScrFI,<br>-BstNI,<br>+MaeI,<br>+BfaI |

|                             |   |    |                 |            |         |      |  |
|-----------------------------|---|----|-----------------|------------|---------|------|--|
| 5122/5123de<br>ICA          | del of CA<br>at<br>5122/5123,<br>A to T<br>at 5121  | 43 | Frameshift      | Japanese   | MM14    | Het  | no change                                |
| R1586X                      | CGA to TGA<br>at 5129   | 43 | Stop at<br>1586 | Japanese   | MM12    | Hom  | +Dde I                                   |
| 5245delG                    | del of G<br>at 5245<br>and G to<br>C at 5249,<br>or G to C<br>at 5245<br>and del G<br>at 5249 | 44 | Frameshift      | French     | MM63    | Hom  | -Bpm I,<br>-BanII<br>+ AvaII,<br>+Sau96I |
| 5 E1732X                    | GAG to TAG<br>at 5567   | 46 | Stop at<br>1732 | Spanish    | MM73    | Het  | -Mbo II                                  |
| 2573-77<br>Hom<br>del ACCCA | Del of ACCCA at 23<br>?Please provide<br>2573-77  | 23 |                 | Frameshift | Italian | MM69 |  |

<sup>1</sup> MM: Miyoshi myopathy; DMAT: distal myopathy with anterior tibial onset; LGMD: limb girdle muscular dystrophy

<sup>2</sup> +: create a new restriction site, -: eliminate an existing restriction site.

- 30 -

Twelve of the eighteen different mutations are predicted to block dysferlin expression, either through nonsense or frameshift changes. Seven of the thirteen samples are homozygous and thus expected to result in complete loss of dysferlin function. For each mutated exon in these patients, at least 50 control DNA samples (100 chromosomes) were screened to determine the frequencies of the sequence variants. When possible, the parents and siblings of affected individuals were also screened to verify that defined mutations were appropriately co-inherited with the disease in each pedigree (Fig. 4). In two families (50, 58 in Table 2) heterozygous mutations were identified in one allele (respectively a missense mutation and a 2 bp deletion). Mutations in the other allele are presumed to have not been detected (or in three of the screened MM families) either because the mutant and normal SSCP products are indistinguishable or because the mutation lies outside of coding sequence (i.e., in the promoter or a regulatory region of an intron). The disease-associated mutations did not appear to arise in the population as common polymorphisms.

More mutations can be identified by using appropriate primer pairs to amplify an exon and analyze its sequence. The following primer pairs are useful for exon amplification.

| Exon Code | Primer Sequence                                 |
|-----------|---|
| 1 F408    | 5'-gaccacaaagcggcgcctcgg-3' {SEQ ID NO: 130}    |
| F4101     | 5'-gaccccggcgaggggtggtcgg-3' {SEQ ID NO: 131}   |
| 2 F4111   | 5'-tgtctctccattctcccttttg-3' {SEQ ID NO: 132}   |
| R4111     | 5'-aggacactgctgagaaggcacctc-3' {SEQ ID NO: 133} |

- 31 -

|                |    |       |  |
|----------------|----|-------|--|
|                | 3  | F3262 | 5-agtgccctggtggcacgaagg-3' {SEQ ID     |
| NO: 134}       |    |       |  |
|                |    | R3261 | 5-cctacctgcaccttcaagccatgg-3' {SEQ ID  |
| NO: 135}       |    |       |  |
| 5              | 4  | F3251 | 5-cagaagagccaggggtgccttagg-3' {SEQ ID  |
| NO: 136}       |    |       |  |
|                |    | R3251 | 5-ccttggaccttaacctggcagagg-3' {SEQ ID  |
| NO: 137}       |    |       |  |
|                | 5  | F3242 | 5-cgaggccagcgcaccaacctg-3' {SEQ ID     |
| 10 NO: 138}    |    |       |  |
|                |    | R3242 | 5-actgccggccattcttgcctggg-3' {SEQ ID   |
| NO: 139}       |    |       |  |
|                | 6  | F3231 | 5-ccaggcctcattagggccctc-3' {SEQ ID     |
| NO: 140}       |    |       |  |
| 15             |    | R3231 | 5-ctgaagaggagcctgggggtcag-3' {SEQ ID   |
| NO: 141}       |    |       |  |
|                | 7  | F3222 | 5-ctgagatttctgactcttgggggtg-3' {SEQ ID |
| NO: 142}       |    |       |  |
|                |    | R3211 | 5-aaggttctgccctcatgccccatg-3' {SEQ ID  |
| 20 NO: 143}    |    |       |  |
|                | 8  | F3561 | 5-ctggcctgagggatcagcagg-3' {SEQ ID     |
| NO: 144}       |    |       |  |
|                |    | R3561 | 5-gtgcatacatacagcccacggag-3' {SEQ ID   |
| NO: 145}       |    |       |  |
| 25             | 9  | F3551 | 5-gagctattgggttggccgtgtggg-3' {SEQ ID  |
| NO: 146}       |    |       |  |
|                |    | R3552 | 5-accaacacggagaagtgagaactg-3' {SEQ ID  |
| NO: 147}       |    |       |  |
|                | 10 | F3201 | 5-ccacactttattttaacgctttggcgg-3' {SEQ  |
| 30 ID NO: 148} |    |       |  |
|                |    | R3201 | 5-cagaaccaaaaatgcaaggatacgg-3' {SEQ ID |
| NO: 149}       |    |       |  |
|                | 11 | F3191 | 5-cttctgattctgggatcaccaaagg-3' {SEQ    |
| ID NO: 150}    |    |       |  |

- 32 -

|             |          |   |
|-------------|----------|---|
|             | F3191    | 5-ggaccgtaaggaagacccaggg-3' {SEQ ID     |
| NO: 151}    |          |   |
| 12          | F3181    | 5-cctgtgctcaggagcgcacatgaagg-3' {SEQ ID |
| NO: 152}    |          |   |
| 5           | R3181    | 5-gcagacctcccacccaagggcg-3' {SEQ ID     |
| NO: 153}    |          |   |
| 13          | F3171    | 5-gagacagatgggggacagtcaggg-3' {SEQ ID   |
| NO: 154}    |          |   |
|             | R3171    | 5-cctcccgagagaaccctcctg-3' {SEQ ID      |
| 10 NO: 155} |          |   |
| 14          | F3161    | 5-gggagcccagagtccccatgg-3' {SEQ ID      |
| NO: 156}    |          |   |
|             | R3161    | 5-gggcctccttgggtttgctgg-3' {SEQ ID      |
| NO: 157}    |          |   |
| 15          | 15 F3541 | 5-gcctccccagcatcctgccgg-3' {SEQ ID      |
| NO: 158}    |          |   |
|             | R3541    | 5-tcactgagccgaatgaaactgagg-3' {SEQ      |
| ID NO: 159} |          |   |
| 16          | F3531    | 5-tgtggcctgagttcctttcctgtg-3' {SEQ ID   |
| 20 NO: 160} |          |   |
|             | R3531    | 5-ggtcaaagggcagaacgaagaggg-3' {SEQ ID   |
| NO: 161}    |          |   |
| 17          | F3151    | 5-cccgctccttctcccagccatg-3' {SEQ ID     |
| NO: 162}    |          |   |
| 25          | R3151    | 5-ctccccctggttgtccccaagg-3' {SEQ ID     |
| NO: 163}    |          |   |
| 18          | F3141    | 5-cgaccctctgattgccacttggtg-3' {SEQ ID   |
| NO: 164}    |          |   |
|             | R3141    | 5-ggcatcctgcccttgccaggg-3' {SEQ ID      |
| 30 NO: 165} |          |   |
| 19          | F3522    | 5-tctgtctccccctgctccttg-3' {SEQ ID NO:  |
| 166}        |          |   |
|             | R3522    | 5-cttccctgccccgacgccag-3' {SEQ ID       |
| NO: 167}    |          |   |



- 33 -

|    |             |       |  |
|----|-------------|-------|--|
|    | 20          | F3121 | 5-gctcctcccgtgaccctctgg-3' {SEQ ID     |
|    | NO: 103}    |       |  |
|    |             | R3121 | 5-gggtcccagccaggagcactg-3' {SEQ ID     |
|    | NO: 104}    |       |  |
| 5  | 21          | F3111 | 5-cagcgctcaggcccgtctctc-3' {SEQ ID     |
|    | NO: 168}    |       |  |
|    |             | R3111 | 5-tgcataggcatgtgcagctttggg-3' {SEQ ID  |
|    | NO: 169}    |       |  |
|    | 22          | F3512 | 5-catgcaccctctgccctgtgg-3' {SEQ ID     |
| 10 | NO: 170}    |       |  |
|    |             | R3512 | 5-agttgagccaggagaggtggg-3' {SEQ ID     |
|    | NO: 171}    |       |  |
|    | 23          | F3101 | 5-catcaggcgcatcctcatctgtccg-3' {SEQ ID |
|    | NO: 172}    |       |  |
| 15 |             | R3091 | 5-agcaggagagcagaagaagaaagg-3' {SEQ ID  |
|    | NO: 173}    |       |  |
|    | 24          | F3082 | 5-gtgtgtcaccatccccaccccg-3' {SEQ ID    |
|    | NO: 174}    |       |  |
|    |             | R3082 | 5-caagagatgggagaaaggccttatg-3' {SEQ    |
| 20 | ID NO:175}  |       |  |
|    | 25          | F3073 | 5-ctgggacatccggatcctgaagg-3' {SEQ ID   |
|    | NO: 176}    |       |  |
|    |             | R3073 | 5-tccaggtagtgggaggcagagg-3' {SEQ ID    |
|    | NO: 177}    |       |  |
| 25 | 26          | F3061 | 5-tcccactacctggagctgccttgg-3' {SEQ     |
|    | ID NO: 178} |       |  |
|    |             | R3051 | 5-ggctctccccagccctccctg-3' {SEQ ID     |
|    | NO: 179}    |       |  |
|    | 27          | F3601 | 5-cagagcagcagagactctgaccag-3' {SEQ     |
| 30 | ID NO: 180} |       |  |
|    |             | R3601 | 5-tagaccccacctgccctgag-3' {SEQ ID      |
|    | NO: 181}    |       |  |
|    | 28          | F3501 | 5-tcctctcattgcttgctgttcgg-3' {SEQ      |
|    | ID NO: 182} |       |  |

- 34 -

|                |       |                                |         |
|----------------|-------|--------------------------------|---------|
|                | R3501 | 5-ttgagagcttgccggggatgg-3'     | {SEQ ID |
| NO: 183}       |       |                                |         |
| 29             | F3031 | 5-aagtgcccaagcaatgagtgaccgg-3' | {SEQ    |
| ID NO: 184}    |       |                                |         |
| 5              | R3021 | 5-ctcactcccacccaccacctg-3'     | {SEQ ID |
| NO: 185}       |       |                                |         |
| 30             | F3011 | 5-cccaccggcctctgagttctgc-3'    | {SEQ ID |
| NO: 186}       |       |                                |         |
|                | R3001 | 5-accctacccaagccaggacaagtg-3'  | {SEQ    |
| 10 ID NO: 187} |       |                                |         |
| 31             | F2141 | 5-gaatctgccataaccagcttcgtg-3'  | {SEQ    |
| ID NO: 188}    |       |                                |         |
|                | R2141 | 5-tatcaccccatagaggcctcgaag-3'  | {SEQ    |
| ID NO: 189}    |       |                                |         |
| 15 32          | F2981 | 5-cagccactcactctggcacctctg-3'  | {SEQ    |
| ID NO: 190}    |       |                                |         |
|                | R2981 | 5-agcccacagtctctgactctcctg-3'  | {SEQ    |
| ID NO: 191}    |       |                                |         |
| 33             | F2131 | 5-acatctctcagggtccttgctgtg-3'  | {SEQ    |
| 20 ID NO: 192} |       |                                |         |
|                | R2211 | 5-cctgtgaggggacgaggcagg-3'     | {SEQ ID |
| NO: 193}       |       |                                |         |
| 34             | F2202 | 5-gccctgggtaagggatgctgattc-3'  | {SEQ    |
| ID NO: 194}    |       |                                |         |
| 25             | R2202 | 5-cctgcctgggcctcctggatc-3'     | {SEQ ID |
| NO: 195}       |       |                                |         |
| 35             | F2111 | 5-gaggggtgatgggggccttagg-3'    | {SEQ ID |
| NO: 196}       |       |                                |         |
|                | R2112 | 5-gcaatcagtttgaagaaggaaagg-3'  | {SEQ    |
| 30 ID NO: 197} |       |                                |         |
| 36             | F2102 | 5-cccctctcaccatctcctgatgtg-3'  | {SEQ    |
| ID NO: 105}    |       |                                |         |
|                | R2111 | 5-ggcttcaccttcctctacctcgg-3'   | {SEQ    |
| ID NO: 106}    |       |                                |         |

- 35 -

|    |             |       |                                     |
|----|-------------|-------|-------------------------------------|
|    | 37          | F2101 | 5-cacctttgtctccattctacctgc-3' {SEQ  |
|    | ID NO: 198} |       |                                     |
|    |             | R2101 | 5-ctcccagccccccacgcccagg-3' {SEQ ID |
|    | NO: 199}    |       |                                     |
| 5  | 38          | F2091 | 5-ctgagccactctcctcattctgtg-3' {SEQ  |
|    | ID NO: 200} |       |                                     |
|    |             | R2091 | 5-tggaaggggacagtagggagg-3' {SEQ ID  |
|    | NO: 201}    |       |                                     |
|    | 39          | F2081 | 5-ggccagtgcgttcttcctcctc-3' {SEQ ID |
| 10 | NO: 202}    |       |                                     |
|    |             | R2071 | 5-tccctgacctgcccacatctc-3' {SEQ ID  |
|    | NO: 203}    |       |                                     |
|    | 40          | F2061 | 5-gcccctgtcaggcctggatgg-3' {SEQ ID  |
|    | NO: 204}    |       |                                     |
| 15 |             | R2061 | 5-tgaccagggcctccctggagg-3' {SEQ ID  |
|    | NO: 205}    |       |                                     |
|    | 41          | F2051 | 5-ctgaaatgggtctctttctttctac-3' {SEQ |
|    | ID NO: 206} |       |                                     |
|    |             | R2051 | 5-cacaccgactgtcagactgaagag-3' {SEQ  |
| 20 | ID NO: 207} |       |                                     |
|    | 42          | F2041 | 5-ttgtcccctcctctaatacccatg-3' {SEQ  |
|    | ID NO: 208} |       |                                     |
|    |             | R2041 | 5-ggggttagggacgtcttcgagg-3' {SEQ ID |
|    | NO: 209}    |       |                                     |
| 25 | 43          | F2031 | 5-cagccaaaccatatcaacaatg-3' {SEQ ID |
|    | NO: 210}    |       |                                     |
|    |             | R2021 | 5-ctggggagggtgagggtcttag-3' {SEQ ID |
|    | NO: 211}    |       |                                     |
|    | 44          | F2011 | 5-gaagtgttttgtctcctcctc-3' {SEQ ID  |
| 30 | NO: 212}    |       |                                     |
|    |             | R2011 | 5-gcaggcagccagcccccatc-3' {SEQ ID   |
|    | NO: 213}    |       |                                     |
|    | 45          | F1021 | 5-ggggtgccctgtgttggtgac-3' {SEQ ID  |
|    | NO: 214}    |       |                                     |

- 36 -

R1031 5-gcaggcagccagcccccatc-3' {SEQ ID  
 NO: 215}  
 46 F1041 5-ctcgtctatgtcttgtgcttgctc-3' {SEQ  
 ID NO: 216}  
 5 R1051 5-caccatgggtttggggcatgtgg-3' {SEQ ID  
 NO: 217}  
 47 F1061 5-tctcgcttccccagctcctgc-3' {SEQ ID  
 NO: 218}  
 R1061 5-tctggagttcgaggactctggg-3' {SEQ ID  
 10 NO: 219}  
 48 F1071 5-agaaggggtggggagagaaacgg-3' {SEQ ID  
 NO: 220}  
 R1071 5-cagctcagagcctgtggctgg-3' {SEQ ID  
 NO: 221}  
 15 49 F1082 5-aaggccttcccatcctttggtagg-3' {SEQ  
 ID NO: 222}  
 R1082 5-acaaccagagggagcacggg-3' {SEQ ID  
 NO: 223}  
 50 F1092 5-gttgacgatgtatataactgtgttg-3' {SEQ  
 20 ID NO: 224}  
 R1091 5-gctggcaccacagggaatcgg-3' {SEQ ID  
 NO: 110}  
 51 F1102 5-gcctctctctaactttgcttccttg-3' {SEQ  
 ID NO: 225}  
 25 R1101 5-agcccccatgtgcagaatggg-3' {SEQ ID  
 NO: 112}  
 52 F1112 5-ggctacaggctggcagtgatcgag-3' {SEQ  
 ID NO: 226}  
 R1112 5-ttcccccatgccctccactgg-3' {SEQ ID  
 30 NO: 227}  
 53 F1121 5-agccttcgtgcccttaaccaagtg-3' {SEQ  
 ID NO: 228}  
 R1121 5-ctgtgggcattggggctcagg-3' {SEQ ID  
 NO: 229}

- 37 -

54 F1141 5-ggatgccagttgactccggg-3' {SEQ ID  
NO: 115}  
R1141 5-ccccaccacagtgtcgtcagg-3' {SEQ ID  
NO: 116}  
5 55 F1151 5-gccccagtgggatcaccatg-3' {SEQ ID  
NO: 230}  
R116 5-atgctggaggggacccacgg-3' {SEQ ID  
NO: 231}

#### Comparison of Dysferlin With Other Proteins

10 The 6,243 bp ORF of this candidate MM gene is  
predicted to encode 2,080 amino acids (Figs. 1C and 2;  
SEQ ID NO:2). At the amino acid level, this protein is  
highly homologous to the nematode (*Caenorhabditis*  
*elegans*) protein fer-1 (27% identical, 57% identical or  
15 similar: the sequence alignment and comparison was  
performed using [http://vega.igh.cnrs.fr/bin/nph-](http://vega.igh.cnrs.fr/bin/nph-align_query.pl)  
[align\\_query.pl.](http://vega.igh.cnrs.fr/bin/nph-align_query.pl)) (Argon & Ward, 1980, *Genetics* 96:413-33;  
Achanzar & Ward, 1997, *J. Cell Science* 110:1073-81).  
This dystrophy-associated, fer-1-like protein has  
20 therefore been designated "dysferlin."

The fer-1 protein was originally identified through  
molecular genetic analysis of a class of fertilization-  
defective *C. elegans* mutants in which spermatogenesis is  
abnormal (Argon & Ward, 1980, *Genetics* 96:413-33). The  
25 mutant fer-1 spermatozoa have defective mobility and show  
imperfect fusion of membranous organelles (Ward et al.,  
1981, *J. Cell Bio.* 91:26-44). Like fer-1, dysferlin is a  
large protein with an extensive, highly charged  
hydrophilic region and a single predicted membrane  
30 spanning region at the carboxy terminus (Fig. 3). There  
is a membrane retention sequence 3' to the membrane  
spanning stretch, indicating that the protein may be  
preferentially targeted to either endoplasmic or  
sarcoplasmic reticulum, probably as a Type II protein

- 38 -

(i.e. with the NH<sub>2</sub> end and most of the following protein located within the cytoplasm) (Fig. 1C). Several nuclear membrane targeting sequences are predicted within the cytoplasmic domain of the protein

5 (<http://psort.nibb.ac.jp/form.html>). Immunocytochemical detection of dysferlin suggests that dysferlin is targeted to or anchored within the sarcoplasmic reticulum.

The cytoplasmic component of this protein contains  
10 four motifs homologous to C2 domains. C2 domains are intracellular protein modules composed of 80 - 130 amino acids (Rizo & Sudhof, 1998, *J. Biol. Chem.* 273:15897). Originally identified within a calcium-dependent isoform of protein kinase C (Nishizuka, 1988, *Nature* 334:661-65),  
15 C2 domains are present in numerous proteins. These domains often arise in approximately homologous pairs described as double C2 or DOC2 domains. One DOC2 protein, DOC2 $\alpha$ , is brain specific and highly concentrated in synaptic vesicles (Orita et al., 1995, *Biochem.*  
20 *Biophys. Res. Comm.* 206:439-48), while another, DOC2 $\beta$ , is ubiquitously expressed (Sakaguchi et al., 1995, *Biochem. Biophys. Res. Comm.* 217:1053-61). Many C2 modules can fold to bind calcium, thereby initiating signaling events such as phospholipid binding. At distal nerve  
25 terminals, for example, the synaptic vesicle protein synaptotagmin has two C2 domains that, upon binding calcium, permit this protein to interact with syntaxin, triggering vesicle fusion with the distal membrane and neurotransmitter release (Sudhof & Rizo, 1996, *Neuron*  
30 17:379-88).

The four dysferlin C2 domains are located at amino acid positions 32-82, 431-475, 1160-1241, and 1582-1660 (Figs. 1C and 3). Indeed, it is almost exclusively through these regions that dysferlin has homology to any  
35 proteins other than fer-1. Each of these segments in

- 39 -

dysferlin is considerably smaller than a typical C2 domain. Moreover, these segments are more widely separated in comparison with the paired C2 regions in synaptotagmin, DOC2 $\alpha$  and  $\beta$  and related C2-positive proteins. For this reason, it is difficult to predict whether the four relatively short C2 domains in dysferlin function analogously to conventional C2 modules. That dysferlin might, by analogy with synaptotagmin, signal events such as membrane fusion is suggested by the fact that fer-1 deficient worms show defective membrane organelle fusion within spermatozoa (Ward et al., 1981, *J. Cell Bio.* 91:26-44).

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

#### EXAMPLES

##### Example 1: Production of dysferlin protein

Standard methods can be used to synthesize either wild type or mutant dysferlin, or fragments of either. These methods can also be used to synthesize brain-specific dysferlin polypeptides including full-length or fragments (e.g., a polypeptide unique to brain-specific dysferlin). For example, a recombinant expression vector encoding dysferlin (or a fragment thereof: e.g., dysferlin minus its membrane-spanning region) operably linked to appropriate expression control sequences can be used to express dysferlin in a prokaryotic (e.g., *E. coli*) or eukaryotic host (e.g., insect cells, yeast cells, or mammalian cells). The protein is then purified by standard techniques. If desired, DNA encoding part or all of the dysferlin sequence can be joined in-frame to DNA encoding a different polypeptide, to produce a chimeric DNA that encodes a hybrid polypeptide. This can be used, for example, to add a tag that will simplify identification or purification of the expressed protein,

- 40 -

or to render the dysferlin (or fragment thereof) more immunogenic.

The preferred means for making short peptide fragments of dysferlin is by chemical synthesis. These  
5 fragments, like dysferlin itself, can be used to generate antibodies, or as positive controls for antibody-based assays.

Fusion proteins are useful, e.g., for generating antibodies. Such fusion proteins are generated using  
10 known methods. In one example, to construct glutathione S-transferase (GST):dysferlin fusion proteins, the BLAST program (Altschul et al., 1990, J. Molec. Biol. 215:403-410) was used to identify three regions of the dysferlin cDNA that show no homology to any known human proteins  
15 (Figure 1). These were subcloned from the dysferlin cDNA as BstYI (881-1333), XmnI (1990-2718) and SalI (5364-5732) fragments ligated respectively into BamHI, SmaI and SalI sites of pGEX-5X-3 (Pharmacia). The three fragments correspond to amino acid sequences at amino acid  
20 locations 253-403, 624-865, and 1664-1786 of SEQ ID NO:2, respectively. The resulting GST fusion proteins of BamHI (43 kDa) and SmaI (53.3 kDa) formed insoluble aggregates that were isolated by SDS-PAGE. The fusion protein of SalI (40.2 kDa) was soluble and thus could be purified  
25 using a glutathione Sepharose 4B column; the SalI dysferlin fragment (14.2 kDa) was isolated by cleavage from GST using Factor Xa protease. The eluted protein was concentrated and further purified by SDS-PAGE. For all three of the fusion peptides, the resulting SDS-PAGE  
30 bands were excised and used to immunize rabbits.

#### Example 2: Production and characterization of anti-dysferlin antibodies

Techniques for generating both monoclonal and polyclonal antibodies specific for a particular protein



- 41 -

are well known. The antibodies can be raised against a short peptide epitope of dysferlin, an epitope linked to a known immunogen to enhance immunogenicity, a long fragment of dysferlin, or the intact protein. Antibodies  
5 can also be raised against brain-specific dysferlin polypeptides, e.g., against amino acids 1-24 of SEQ ID NO:233. Such antibodies raised against dysferlin or brain-specific dysferlin polypeptides are useful for e.g., localizing such polypeptides in tissue sections or  
10 fractionated cell preparations and diagnosing dysferlin-related disorders.

An isolated dysferlin protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind dysferlin using standard techniques  
15 for polyclonal and monoclonal antibody preparation. The dysferlin immunogen can also be a mutant dysferlin or a fragment of a mutant dysferlin. A full-length dysferlin protein can be used or, alternatively, antigenic peptide fragments of dysferlin can be used as immunogens. The  
20 antigenic peptide of dysferlin comprises at least 8 (preferably 10, 15, 20, or 30) amino acid residues of the amino acid sequence shown in SEQ ID NO:2 and encompasses an epitope of such that an antibody raised against the peptide forms a specific immune complex with dysferlin.  
25 Preferred epitopes encompassed by the antigenic peptide are regions of dysferlin that are located on the surface of the protein, e.g., hydrophilic regions.

A dysferlin immunogen typically is used to prepare antibodies by immunizing a suitable subject (e.g.,  
30 rabbit, goat, mouse or other mammal) with the immunogen. An appropriate immunogenic preparation can contain, for example, recombinantly expressed dysferlin protein or a chemically synthesized dysferlin polypeptide. The preparation can further include an adjuvant, such as  
35 Freund's complete or incomplete adjuvant, or similar

- 42 -

immunostimulatory agent. Immunization of a suitable subject with an immunogenic dysferlin preparation induces a polyclonal anti-dysferlin antibody response.

Polyclonal anti-dysferlin antibodies ("dysferlin  
5 antibodies") can be prepared as described above by immunizing a suitable subject with a dysferlin immunogen. The dysferlin antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using  
10 immobilized dysferlin. If desired, the antibody molecules directed against dysferlin can be isolated from the mammal (e.g., from the blood) and further purified by well-known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after  
15 immunization, e.g., when the dysferlin antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein  
20 (1975) *Nature* 256:495-497, the human B cell hybridoma technique (Kozbor et al. (1983) *Immunol. Today* 4:72), the EBV-hybridoma technique (Cole et al. (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for  
25 producing hybridomas is well known (see generally *Current Protocols in Immunology* (1994) Coligan et al. (eds.) John Wiley & Sons, Inc., New York, NY). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with a  
30 dysferlin immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds dysferlin.

Any of the many well known protocols used for fusing  
35 lymphocytes and immortalized cell lines can be applied

- 43 -

for the purpose of generating a monoclonal antibody against dysferlin (see, e.g., *Current Protocols in Immunology*, supra; Galfre et al. (1977) *Nature* 266:55052; R.H. Kenneth, in *Monoclonal Antibodies: A New Dimension* 5 *In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); and Lerner (1981) *Yale J. Biol. Med.*, 54:387-402. Moreover, the one in the art will appreciate that there are many variations of such methods which also would be useful. Hybridoma cells producing a 10 monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind dysferlin, e.g., using a standard ELISA assay.

Alternative to preparing monoclonal antibody- 15 secreting hybridomas, a monoclonal dysferlin antibody can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (e.g., an antibody phage display library) with dysferlin to thereby isolate immunoglobulin library members that bind dysferlin. Kits 20 for generating and screening phage display libraries are commercially available (e.g., the Pharmacia Recombinant Phage Antibody System, Catalog No. 27-9400-01; and the Stratagene SurfZAP™ Phage Display Kit, Catalog No. 240612). Additionally, examples of methods and reagents 25 particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs et al. (1991) *Bio/Technology* 9:1370-1372; Hay et al. (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse et al. (1989) *Science*

- 44 -

246:1275-1281; Griffiths et al. (1993) *EMBO J.* 12:725-734.

As an example, two polyclonal antisera were raised for each of the fusion peptide antigens described above using New Zealand White rabbits. The rabbits were injected with 0.5 mg of antigen using keyhole limpet hemocyanin (KLH) as the adjuvant. Booster injections of 0.25 mg antigen were administered every three weeks over 12 weeks. Serum was prepared from the rabbits and was purified using affinity column chromatography (HiTrap; Pharmacia) or antigen-blotted polyvinylidene difluoride (PVDF) membrane.

Immunoblotting was used to verify that the affinity-purified antisera recognize the cognate fusion peptides by Western immunoblotting (WIB) and that this reactivity was immunoadsorbed by pre-incubation of the antisera with the peptides. Thus, antiserum raised against the polypeptide encoded by the SalI fragment (encoding amino acids 1664-1786) identified the fragment both as a cleaved, 14.2 kDa fragment and as a component of the 40.2 kDa GST-SalI fusion peptide. No reactivity was evident in the fraction containing only the GST fusion partner. Immunoadsorption entirely abolished this staining. Analogous results were detected with all six antisera (to the three different target fusion peptides).

#### Preparation of subcellular fractions

Frozen human muscle (0.3 g) was homogenized in five volumes of 0.25 M sucrose containing proteinase inhibitor (Complete, Boehringer). Subcellular fractions of nuclei, mitochondria, microsomes, and cytosol were separated by differential centrifugation. The purity of each fraction was evaluated by immunoblotting of fraction-specific proteins with antibodies to histone H1 (Calbiochem), cytochrome c (Santa Cruz), Na<sup>+</sup>-K<sup>+</sup> ATPase  $\alpha$ 1 subunit

- 45 -

(Research Diagnostics) and cytosolic superoxide dismutase (Calbiochem).

#### Dysferlin in subcellular fractions

Immunoblotting was used to analyze dysferlin  
5 expression. Twenty  $\mu\text{g}$  of each subcellular fraction and  
40  $\mu\text{g}$  of whole homogenate of muscle were separated by  
SDS-PAGE (4-15% gradient gel) and transferred to a  
nitrocellulose membrane. Immunoblotting was performed  
according to standard methods, using chemiluminescence  
10 (ECL, Amersham). Immunoblotting of multi-tissue blots  
identified prominent dysferlin positively at  
approximately 230 kDa in heart, placenta, skeletal muscle  
and kidney. Little or no immuno-positive staining was  
detected in brain, liver, spleen, ovary, or testis.  
15 Lower molecular weight bands (approximately 40 kDa) were  
also evident. Immunoabsorption with the corresponding  
fusion peptide abolished both the large and the smaller  
bands. The 230 kDa band was observed with all of the  
affinity purified, anti-dysferlin antisera.  
20 Immunoblotting of fractionated human muscle  
documented distinct 230 kDa bands in the whole muscle  
homogenate and in microsomal and nuclear fractions. Some  
immunoreactivity was also evident in the nuclear and  
mitochondrial fractions. No immunoreactivity was  
25 detected in the cytosolic fractions. This pattern was  
seen with all of the anti-dysferlin antisera, and was  
eliminated by immunoabsorption. The identity of the  
assayed fractions was verified by Western blotting using  
fraction-specific antibodies: histone H1 for the nuclear  
30 fraction, cytochrome c for the mitochondrial fraction,  
Na<sup>+</sup>-K<sup>+</sup> ATPase  $\alpha$ 1-subunit for the microsomal fraction, and  
SOD1 for the cytosolic fraction.

#### Example 3: Diagnosis

- 46 -

The discovery of mutations in the dysferlin gene that are associated with the MM and LMGD2B phenotypes means that individuals can be tested for the disease gene before symptoms appear. This will permit genetic testing and counseling of those with a family history of the disease. Additionally, individuals diagnosed with the genetic defect can be closely monitored for the appearance of symptoms, thereby permitting early intervention, including genetic therapy, as appropriate. Individuals with a brain-specific dysferlin-related disorder can be diagnosed using such methods.

Diagnosis can be carried out on any suitable genomic DNA sample from the individual to be tested. Typically, a blood sample from an adult or child, or a sample of placental or umbilical cord cells of a newborn would be used; alternatively, one could utilize a fetal sample obtained by amniocentesis or chorionic villi sampling.

It is expected that standard genetic diagnostic methods can be used. For example, PCR can be utilized to identify the presence of a deletion, addition, or substitution of one or more nucleotides within any one of the exons of dysferlin. Following the PCR reaction, the PCR product can be analyzed by methods such as a heteroduplex detection technique based upon that of White et al. (1992, *Genomics* 12:301-06), or by techniques such as cleavage of RNA-DNA hybrids using RNase A (Myers et al., 1985, *Science* 230:1242-46), single-stranded conformation polymorphism (SSCP) analysis (Orita et al., 1989, *Genomics* 10:298-99), di-deoxy-fingerprinting (DDF) (Blaszyk et al., 1995, *Biotechniques* 18: 256-260) and denaturing gradient gel electrophoresis (DGGE; Myers et al., 1987, *Methods Enzymol.* 155:501-27). The PCR may be carried out using a primer which adds a G+C rich sequence (termed a "GC-clamp") to one end of the PCR product, thus improving the sensitivity of the subsequent DGGE

- 47 -

procedure (Sheffield et al., 1989, *Proc. Natl. Acad. Sci. USA* 86:232-36). If the particular mutation present in the patient's family is known to have removed or added a restriction site, or to have significantly increased or  
5 decreased the length of a particular restriction fragment, a protocol based upon restriction fragment length polymorphism (RFLP) analysis (perhaps combined with PCR) may be appropriate.

The apparent genetic heterogeneity resulting in the  
10 MM/LGMD2B phenotypes means that the nature of the particular mutation carried by affected individuals in the patient's family may have to be ascertained prior to attempting genetic diagnosis of the patient. Alternatively, a battery of tests designed to identify  
15 any of several mutations known to result in MM/LGMD2B may be utilized to screen individuals without a defined familial genotype. The analysis can be carried out on any genomic DNA derived from the patient, typically from a blood sample.

20 Instead of basing the diagnosis on analysis of the genomic DNA of a patient, one could seek evidence of the mutation in the level or nature of the relevant expression products. Well-known techniques for analyzing expression include mRNA-based methods, such as Northern  
25 blots and *in situ* hybridization (using a nucleic acid probe derived from the relevant cDNA), and quantitative PCR (as described in St-Jacques et al., 1994, *Endocrinology* 134:2645-57). One could also employ polypeptide based methods, including the use of  
30 antibodies specific for the polypeptide of interest. These techniques permit quantitation of the amount of expression of a given gene in the tissue of interest, at least relative to positive and negative controls. One would expect an individual who is heterozygous for a  
35 genetic defect affecting the level of expression of

- 48 -

dysferlin to show up to a 50% loss of expression of this gene in such a hybridization or antibody-based assay. An antibody specific for the carboxy terminal end would be likely to pick up (by failure to bind to) most or all frameshift and premature termination signal mutations, as well as deletions of the carboxy terminal sequence. Use of a battery of monoclonal antibodies specific for different epitopes of dysferlin would be useful for rapidly screening cells to detect those expressing mutant forms of dysferlin (i.e., cells which bind to some dysferlin-specific monoclonal antibodies, but not to others), or for quantifying the level of dysferlin on the surface of cells. One could also use a protein truncation assay (Heim et al., 1994, *Nature Genetics* 8:218-19) to screen for any genetic defect which results in the production of a truncated polypeptide instead of the wild type protein.

Use of immunodetection to identify normal and disease-associated dysferlin

In the following example, immunodetection methods are used to demonstrate a detectable difference in muscles homogenates between normal and disease-associated dysferlin alleles.

Frozen muscle samples (quadriceps) were homogenized in ten volumes of SDS-PAGE sample buffer and boiled for 5 minutes. The final loading volume of SDS-PAGE was adjusted after densitometric measurements (NIH Image) of myosin heavy chain on the Coomassie blue stained gels. Studies were performed on six MM, two LGMD-2B, and three normal muscle samples.

Immunocytochemistry was performed on 8 micron cryostat sections of the muscle that were fixed in 100% cold acetone for 5 minutes and preincubated with PBS containing 1% BSA, 5% heat-inactivated goat serum and 0.2% Triton®X-100. The sections were incubated with



- 49 -

primary antibodies overnight at 4°C and fluorescein-labeled secondary (TAGO Immunologicals) for 30 minutes at room temperature. The primary antibodies were applied in two double staining combinations: SalI-1 anti-dysferlin  
5 and anti-dystrophin antibodies, and SalI-2 anti-dysferlin and anti- $\delta$ -sarcoglycan antibodies. The sections were mounted in SlowFade (Molecular Probes).

The 230 kDA antigen was absent in samples from all five MM patient in immunoblot assays. All five patients  
10 had normal patterns of dystrophin expression. Genetic analysis of the dysferlin gene in the patients predicted that at least two of the five MM patients should have no full-length protein. Two of the other three patients had mutations in at least one allele that are predicted to  
15 eliminate normal dysferlin expression. In all five patients, absence of dysferlin immuno-staining was documented with at least two other anti-dysferlin anti-sera.

Immunostaining of dysferlin, dystrophin and  $\delta$ -  
20 sarcoglycan proteins demonstrated distinct membrane-associated positivity for each protein in normal muscle. By contrast, in both MM and LGMD-2B muscle the dysferlin protein was absent, while the dystrophin and  $\delta$ -sarcoglycan proteins appeared normal.

## 25 Therapeutic Treatment

A patient with MM/LGMD2B, or an individual genetically susceptible to contracting one or both of these diseases, can be treated by supplying dysferlin therapeutic agents of the present invention. Dysferlin  
30 therapeutic agents include a DNA or a subgenomic polynucleotide coding for a functional dysferlin protein. A DNA (e.g., a cDNA) is prepared which encodes the wild type form of the gene operably linked to expression control elements (e.g., promoter and enhancer) that

- 50 -

induce expression in skeletal muscle cells or any other affected cells. The DNA may be incorporated into a vector appropriate for transforming the cells, such as a retrovirus, adenovirus, or adeno-associated virus. One  
5 of the many other known types of techniques for introducing DNA into cells *in vivo* may be used (e.g., liposomes). Particularly useful would be naked DNA techniques, since naked DNA is known to be readily taken up by skeletal muscle cells upon injection into muscle.  
10 Wildtype dysferlin protein can also be administered to an individual who either expresses mutant dysferlin protein or expresses an inadequate amount of dysferlin protein, e.g., a MM/LGMD2B patient.

Administration of the dysferlin therapeutic agents  
15 of the invention can include local or systemic administration, including injection, oral administration, particle gun, or catheterized administration, and topical administration. Various methods can be used to administer the therapeutic dysferlin composition directly  
20 to a specific site in the body. For example, a specific muscle can be located and the therapeutic dysferlin composition injected several times in several different locations within the body of the muscle. The therapeutic dysferlin composition can be directly  
25 administered to the surface of the muscle, for example, by topical application of the composition. X-ray imaging can be used to assist in certain of the above delivery methods. Combination therapeutic agents, including a dysferlin protein or polypeptide or a subgenomic  
30 dysferlin polynucleotide and other therapeutic agents, can be administered simultaneously or sequentially.

Receptor-mediated targeted delivery of therapeutic compositions containing dysferlin subgenomic polynucleotides to specific tissues can also be used.  
35 Receptor-mediated DNA delivery techniques are described

- 51 -

in, for example, Findeis et al. (1993), *Trends in Biotechnol.* 11, 202-05; Chiou et al. (1994), *Gene Therapeutics: Methods and Applications of Direct Gene Transfer* (J.A. Wolff, ed.); Wu & Wu (1988), *J. Biol. Chem.* 263, 621-24; Wu et al. (1994), *J. Biol. Chem.* 269, 542-46; Zenke et al. (1990), *Proc. Natl. Acad. Sci. U.S.A.* 87, 3655-59; Wu et al. (1991), *J. Biol. Chem.* 266, 338-42.

Alternatively, a dysferlin therapeutic composition can be introduced into human cells *ex vivo*, and the cells then implanted into the human. Cells can be removed from a variety of locations including, for example, from a selected muscle. The removed cells can then be contacted with the dysferlin therapeutic composition utilizing any of the above-described techniques, followed by the return of the cells to the human, preferably to or within the vicinity of a muscle. The above-described methods can additionally comprise the steps of depleting fibroblasts or other contaminating non-muscle cells subsequent to removing muscle cells from a human.

Both the dose of the dysferlin composition and the means of administration can be determined based on the specific qualities of the therapeutic composition, the condition, age, and weight of the patient, the progression of the disease, and other relevant factors. If the composition contains dysferlin protein or polypeptide, effective dosages of the composition are in the range of about 1  $\mu$ g to about 100 mg/kg of patient body weight, e.g., about 50  $\mu$ g to about 50 mg/kg of patient body weight, e.g., about 500  $\mu$ g to about 5 mg/kg of patient body weight.

Therapeutic compositions containing dysferlin subgenomic polynucleotides can be administered in a range of about 0.1  $\mu$ g to about 10 mg of DNA/dose for local administration in a gene therapy protocol. Concentration

- 52 -

ranges of about 0.1  $\mu$ g to about 10 mg, e.g., about 1  $\mu$ g to about 1 mg, e.g., about 10  $\mu$ g to about 100  $\mu$ g of DNA can also be used during a gene therapy protocol. Factors such as method of action and efficacy of transformation and expression are considerations that will effect the dosage required for ultimate efficacy of the dysferlin subgenomic polynucleotides. Where greater expression is desired over a larger area of tissue, larger amounts of dysferlin subgenomic polynucleotides or the same amounts readministered in a successive protocol of administrations, or several administrations to different adjacent or close tissue portions of for example, a muscle site, may be required to effect a positive therapeutic outcome. In all cases, routine experimentation in clinical trials will determine specific ranges for optimal therapeutic effect.

#### Animal Model

A line of transgenic animals (e.g., mice, rats, guinea pigs, hamsters, rabbits, or other mammals) can be produced bearing a transgene encoding a defective form of dysferlin. Standard methods of generating such transgenic animals would be used, e.g., as described below.

Alternatively, standard methods of producing null (i.e., knockout) mice could be used to generate a mouse which bears one defective and one wild type allele encoding dysferlin. If desired, two such heterozygous mice could be crossed to produce offspring which are homozygous for the mutant allele. The homozygous mutant offspring would be expected to have a phenotype comparable to the human MM and/or LGMD2B phenotype, and so serve as models for the human disease.

For example, in one embodiment, dysferlin mutations are introduced into a dysferlin gene of a cell, e.g., a

- 53 -

fertilized oocyte or an embryonic stem cell. Such cells can then be used to create non-human transgenic animals in which exogenous altered (e.g., mutated) dysferlin sequences have been introduced into their genome or

5 homologously recombinant animals in which endogenous dysferlin nucleic acid sequences have been altered. Such animals are useful for studying the function and/or activity of dysferlin and for identifying and/or evaluating modulators of dysferlin function. As used

10 herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep,

15 dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene

20 product in one or more cell types or tissues of the transgenic animal. As used herein, an "homologously recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous dysferlin gene has been altered by homologous

25 recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to completed development of the animal.

A transgenic animal of the invention can be created

30 by introducing a nucleic acid encoding a dysferlin mutation into the male pronuclei of a fertilized oocyte, e.g., by microinjection or retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. A dysferlin cDNA sequence e.g., that of

35 (SEQ ID NO:1 or SEQ ID NO:3) can be introduced as a

- 54 -

transgene into the genome of a non-human animal. Alternatively, a nonhuman homologue of the human dysferlin gene can be isolated based on hybridization to the human dysferlin sequence (e.g., cDNA) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, U.S. Patent No. 4,873,191 and in Hogan, *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the mutant dysferlin transgene in its genome and/or expression of the mutant dysferlin mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene encoding a mutant dysferlin can further be bred to other transgenic animals carrying other transgenes.

To create an homologously recombinant animal, a vector is prepared which contains at least a portion of a dysferlin gene into which a deletion, addition or substitution has been introduced to thereby alter a dysferlin gene. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous dysferlin gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous dysferlin gene is mutated

- 55 -

or otherwise altered (e.g., contains one of the mutations described in Table 2). In the homologous recombination vector, the altered portion of the dysferlin sequence is flanked at its 5' and 3' ends by additional nucleic acid of the dysferlin gene to allow for homologous recombination to occur between the exogenous dysferlin nucleic acid sequence carried by the vector and an endogenous dysferlin gene in an embryonic stem cell. The additional flanking dysferlin nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (see, e.g., Thomas and Capecchi (1987) *Cell* 51:503 for a description of homologous recombination vectors). The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced dysferlin sequence has homologously recombined with the endogenous dysferlin gene are selected (see, e.g., Li et al. (1992) *Cell* 69:915). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see, e.g., Bradley in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley (1991) *Current Opinion in Bio/Technology* 2:823-829 and in PCT Publication Nos. WO 90/11354, WO 91/01140, WO 92/0968, and WO 93/04169.

- 56 -

Other Embodiments

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is  
5 intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.



- 57 -

What is claimed is:

1. An isolated DNA comprising a nucleotide sequence which hybridizes under stringent hybridization conditions to SEQ ID NO:3, or a complement thereof.

5        2. The isolated DNA of claim 1, wherein the nucleotide sequence is SEQ ID NO:117.

3. An isolated DNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:4-12.

4. The isolated DNA of claim 3, comprising the  
10 sequence of SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, or SEQ ID NO:21.

5. An isolated DNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NOS:22-30.

15        6. A single stranded oligonucleotide of 14-50 nucleotides in length having a nucleotide sequence identical to a portion of SEQ ID NO:3, or a complement thereof.

7. A pair of PCR primers consisting of:

20        (a) a first single stranded oligonucleotide consisting of 14-50 contiguous nucleotides that are identical to a portion of SEQ ID NO:117; and

(b) a second single stranded oligonucleotide consisting of 14-50 contiguous nucleotides that are  
25 identical to a portion of SEQ ID NO:117, wherein the sequence of at least one of the oligonucleotides is identical to a portion of a strand of SEQ ID NO:3, and the first oligonucleotide is not complementary to the second oligonucleotide.

- 58 -

8. A pair of single-stranded oligonucleotides,  
wherein both oligonucleotides are selected from the group  
consisting of SEQ ID NOS:130-231, SEQ ID NO:110, and SEQ  
ID NO:112 and the oligonucleotides are different from  
5 each other.

9. An isolated DNA comprising a nucleotide sequence  
that encodes a polypeptide that shares at least 70%  
sequence identity with SEQ ID NO:2, or a complement of  
the nucleotide sequence.

10 10. The isolated DNA of claim 9, wherein the  
polypeptide comprises the sequence of SEQ ID NO:2.

11. An isolated DNA comprising a nucleotide  
sequence which hybridizes under stringent hybridization  
conditions to a nucleic acid having a sequence selected  
15 from the group consisting of SEQ ID NOS:31-79 and 90-100.

12. A single stranded oligonucleotide of 14-50  
nucleotides in length comprising a nucleotide sequence  
which is identical to a portion of a nucleic acid  
selected from the group consisting of SEQ ID NOS:31-79  
20 and 90-100, or a complement of the nucleotide sequence.

13. The oligonucleotide of claim 12, wherein the  
portion includes an intronic sequence.

14. A pair of PCR primers consisting of:  
(a) a first single-stranded oligonucleotide  
25 consisting of 14-50 contiguous nucleotides that are  
identical to a portion of a sense strand of a nucleic  
acid selected from the group consisting of SEQ ID NOS:31-  
85; and

- 59 -

(b) a second single stranded oligonucleotide consisting of 14-50 contiguous nucleotides that are identical to a portion of the antisense strand of a nucleic acid selected from the group consisting of SEQ ID  
5 NOS:31-85, wherein the sequence of at least one of the oligonucleotides comprises a sequence identical to a portion of a nucleic acid selected from SEQ ID NOS: 31-79 and 90-100, and wherein the first oligonucleotide is not complementary to the second oligonucleotide.

10 15. A pair of single-stranded oligonucleotides selected from the group consisting of SEQ ID NOS:101-116, SEQ ID NOS:184-185, SEQ ID NOS:188-191, SEQ ID NOS:210-213, and SEQ ID NOS:216-217.

15 1. 16. A vector comprising the isolated DNA of claim 1.

17. A substantially pure polypeptide comprising an amino acid sequence sharing at least 70% sequence identity with SEQ ID NO:2.

20 18. The substantially pure polypeptide of claim 17, wherein the polypeptide comprises an amino acid sequence identical to that of a naturally occurring polypeptide.

19. The substantially pure polypeptide of claim 18, wherein the amino acid sequence comprises the sequence of SEQ ID NO:2.

25 20. A substantially pure polypeptide comprising an amino acid sequence identical to the amino acid sequence of amino acid residues 1-500, 501-1000, 1001-1500, or 1501-2080 of SEQ ID NO:2.

- 60 -

21. A substantially pure polypeptide comprising the amino acid sequence of SEQ ID NO:86, SEQ ID NO:87, SEQ ID NO:88 or SEQ ID NO:89.

22. A substantially pure polypeptide selected from  
5 the group consisting of amino acids 253-403 of SEQ ID NO:2, amino acids 624-865 of SEQ ID NO:2, and amino acids 1664-1786 of SEQ ID NO:2.

23. A fusion protein comprising a polypeptide of claim 22.

10 24. An antibody that specifically binds to the polypeptide of claim 22.

25. An antibody that binds specifically to the polypeptide of claim 17.

26. A cell comprising the isolated DNA of claim 1.

15 27. A non-human mammal, the genomic DNA of which bears a transgene, wherein the transgene comprises the isolated DNA of claim 1.

28. A transgenic non-human mammal having a transgene disrupting or interfering with the expression  
20 of a dysferlin gene.

29. A method of decreasing the symptoms of muscular dystrophy in a mammal, the method comprising introducing into a cell of said mammal the isolated DNA of claim 1.

30. A method of decreasing the symptoms of muscular  
25 dystrophy in a mammal, the method comprising introducing

- 61 -

into a cell of said mammal the vector of claim 16, the vector being an expression vector.

31. A method of decreasing the symptoms of muscular dystrophy in a mammal, the method comprising introducing  
5 into a cell of said mammal the protein of claim 17.

32. A method for identifying a patient, a fetus, or a pre-embryo at risk for having a dysferlin-related disorder, the method comprising:

(a) obtaining a sample of genomic DNA from the  
10 patient, fetus, or pre-embryo; and

(b) determining whether the sample contains a mutation in a dysferlin gene, wherein a patient, a fetus, or a pre-embryo having a mutation in a dysferlin gene is at risk for having a dysferlin-related disorder.

15 33. The method of claim 32, comprising:

(a) treating the sample of genomic DNA with a restriction enzyme specific for a particular restriction enzyme site; and

(b) detecting the presence or absence of the  
20 particular restriction enzyme site in the sample of genomic DNA as an indication of the presence or absence of a particular mutation in the genomic DNA.

34. The method of claim 33, wherein the restriction enzyme is selected from the group consisting of Pst I,  
25 Fnu4H I, BamH I, BstY I, Ava II, HinP I, Fsp I, Mbo II, ScrF I, BstN I, Mae I, Bfa I, Dde I, Bpm I, Ban II, Ava II, and Sau96 I.

35. The method of claim 32, comprising subjecting the sample to polymerase chain reaction (PCR).

- 62 -

36. The method of claim 32, comprising:

(a) contacting a single stranded oligonucleotide  
with the sample of genomic DNA; and

(c) detecting hybridization or lack thereof between  
5 the single stranded oligonucleotide and the genomic DNA,  
as an indication of the presence or absence of a mutation  
in the genomic DNA.

37. A method for identifying a patient, a fetus, or  
a pre-embryo at risk for having a dysferlin-related  
10 disorder, said method comprising:

(a) providing a sample comprising dysferlin mRNA  
from the patient, fetus, or pre-embryo; and

(b) determining whether the dysferlin mRNA contains  
a mutation, wherein a patient, a fetus, or a pre-embryo  
15 having a dysferlin mRNA containing a mutation is at risk  
for having a dysferlin-related disorder.

38. The method of claim 37, wherein the presence or  
absence of the mutation is detected by Northern blot.

39. The method of claim 37, wherein the method  
20 includes the step of subjecting the sample to polymerase  
chain reaction (PCR).

40. A method for detecting the absence of a  
mutation in a dysferlin protein of a patient, a fetus, or  
a pre-embryo, the method comprising:

25 (a) providing a sample comprising a dysferlin  
protein of the patient, fetus, or pre-embryo;

(b) contacting the sample with the antibody of  
claim 22; and

(c) detecting binding of the antibody to dysferlin  
30 protein in the sample, if any, wherein binding indicates  
a normal dysferlin protein.

- 63 -

41. An isolated DNA comprising a nucleotide sequence that is identical to the sequence of amino acid residues 3501-3520 of SEQ ID NO:1, 3737-3756 of SEQ ID NO:1, 3842-3861 of SEQ ID NO:1, 5114-5139 of SEQ ID NO:1, or 5239-5 5255 of SEQ ID NO:1.

42. An isolated DNA comprising a nucleotide sequence selected from the group consisting of  
3501-3520 of SEQ ID NO:1, wherein nucleotide G at 3510 is A;  
10 3737-3756 of SEQ ID NO:1, wherein nucleotide G at 3746 is deleted;  
3842-3861 of SEQ ID NO:1, wherein nucleotide C at 3851 is T;  
5114-5139 of SEQ ID NO:1, wherein nucleotide C at 15 5122 and nucleotide A at 5123 are deleted;  
5239-5255 of SEQ ID NO:1, wherein nucleotide G at 5245 is deleted and nucleotide G at 5249 is C; and  
5239-5255 of SEQ ID NO:1, wherein nucleotide G at 5245 is C and nucleotide G at 5249 is deleted.

20 43. An isolated nucleic acid comprising a nucleotide sequence which hybridizes under stringent hybridization conditions to nucleic acids 3284-3720 of SEQ ID NO:232, or the complement of said nucleotide sequence.

25 44. An isolated nucleic acid comprising a nucleotide sequence identical to the sequence of nucleotides 3284-3720 of SEQ ID NO:232, or a complement of said nucleotide sequence.

30 45. The isolated nucleic acid of claim 44, wherein the nucleotide sequence comprises the sequence of SEQ ID NO:232 or the complement of SEQ ID NO:232.

- 64 -

46. An isolated polypeptide comprising:
- a) at least 15 contiguous amino acids of the polypeptide comprising amino acids 1-24 of SEQ ID NO:233,
  - b) a naturally occurring allelic variant of a polypeptide comprising amino acids 1-24 of SEQ ID NO:233, or
  - c) an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes under stringent conditions to nucleotides 3284-3720 of SEQ ID NO:232.
- 10 47. The polypeptide of claim 46, wherein the polypeptide comprises SEQ ID NO:233.
48. A vector comprising the nucleic acid of claim 44.
49. A cell comprising the vector of claim 48.
- 15 50. A method of making a polypeptide, the method comprising culturing the cell of claim 49.
51. An antibody which specifically binds to a polypeptide of claim 46.
- 20 52. The antibody of claim 51, wherein the antibody binds to a polypeptide selected from the group comprising amino acids 253-403 of SEQ ID NO:233, amino acids 624-865 of SEQ ID NO:233, and amino acids 1664-1786 of SEQ ID NO:233.
- 25 53. The antibody of claim 51, wherein the antibody is a monoclonal antibody.
54. The antibody of claim 51, wherein the antibody is a polyclonal antibody.



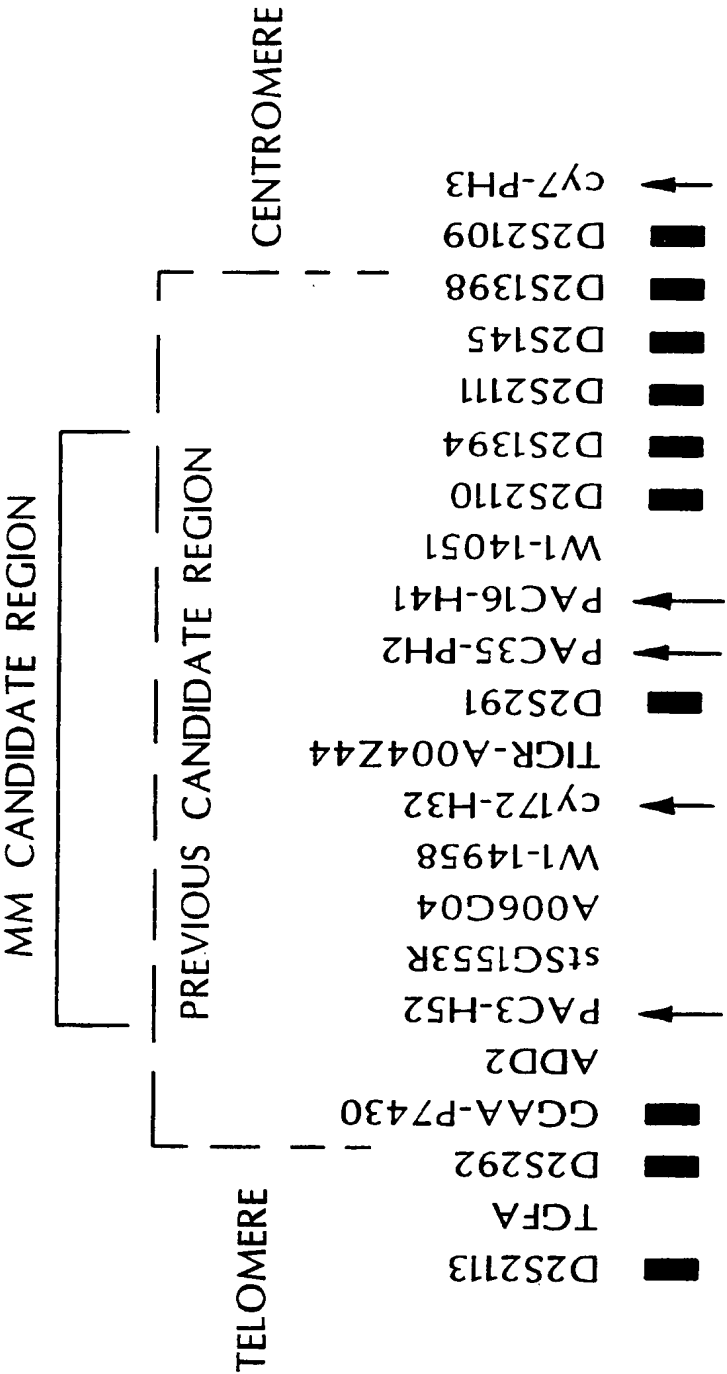
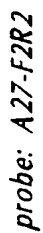
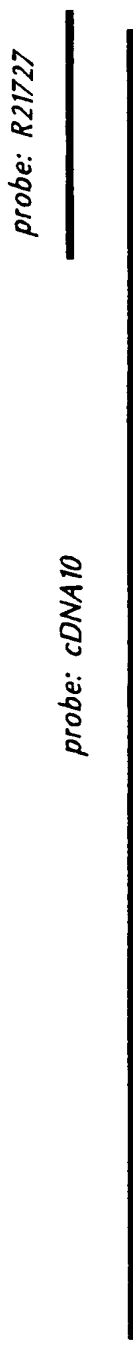
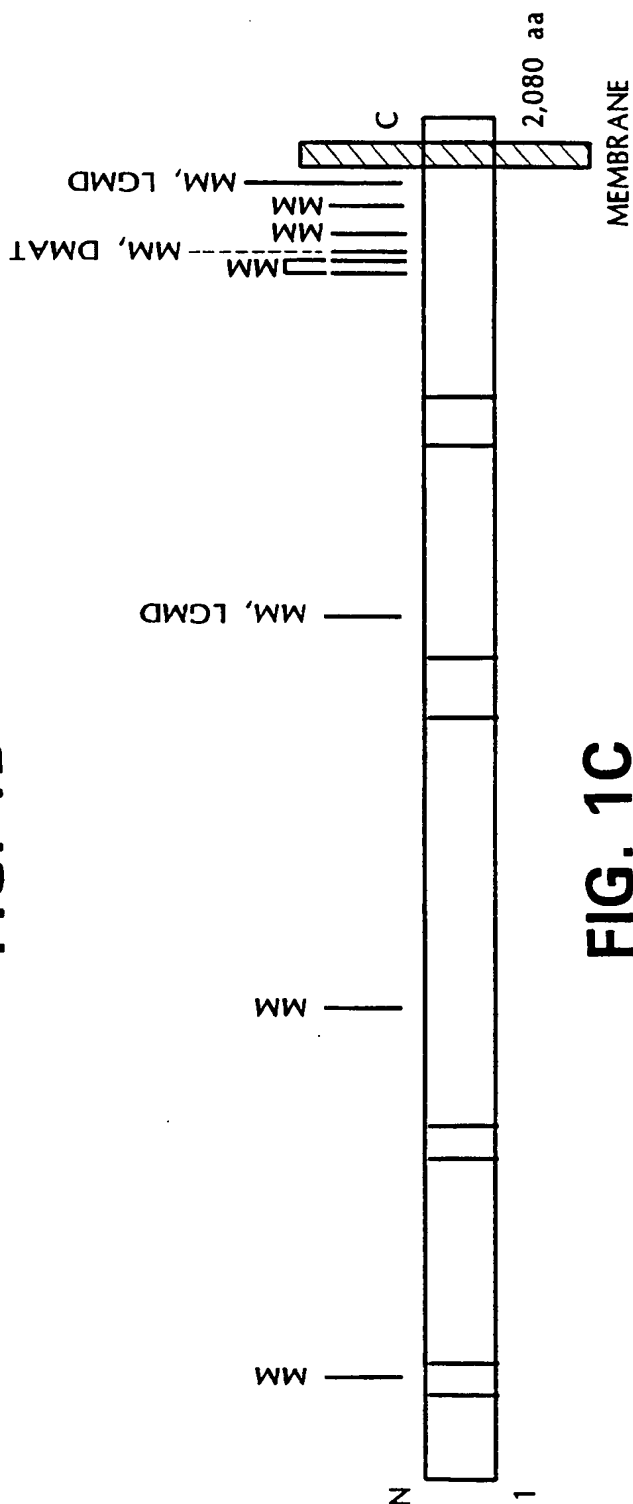


FIG. 1A

2/10



**FIG. 1B**



**FIG. 1C**

3/10

1 ~~MEPWEELVAE~~ NWHTPDTDIS DAYCSAVFAG ~~VKKZTH/TKV~~ ~~SVNPVWNEGE~~  
 51 ~~EDDLKGIPLD~~ ~~CGSELETA/VK~~ ~~DEETMGPNRF~~ ~~LGEAKVPLRE~~ VLATPSLSAS  
 101 FNAPLLDTKK QPTGASLVLO VSYTPLPGAV PLFPFPTPLE PSPTLPDLDO  
 151 VADTGGEDDT EDQGLTGDEA EPFLDQSGGP GAPTTFFKLP ~~SRPPPHYPGE~~  
 201 ~~KKRSAPTST~~ KLLSDKPODF CIRVQVIEGR QLPGVNIXPV VKVTAAGQTK  
 251 RTRIKKGNST LFNELFFNL FDSFGELFDE PIFITVDSR SLRTDALLGE  
 301 FRMDVGTIYR EPRHAYLRKW LLLSDPDDFS AGARGYLKTS LCVLGPDEA  
 351 PLERKDPSED KEDIESNLLR PTGVALRGAN FCLKVFRAD LQOMDDAVMD  
 401 NVKQIFGFES NKGNLVDPFV EVSFAGKMLC ~~SKILEXTANP~~ ~~OWNONITLPA~~  
 451 ~~MEPSMCEKMG~~ ~~LEEDWDDEL~~ ~~ENDIVATTYL~~ SMSKISAPGG EIEEEPAGAV  
 501 KPSKASDLDD YLGFLPTFGP CYINLYGSPR EFTGFPDPYT ELNTGKGEGV  
 551 AYGRLLLSL ETKLVEHSEQ KVEDLPADDI LRVEKYLRRR KYSLFAAFYS  
 601 ATMLQDVDDA IQFEVSIGNY GNKFDMTCLP LASTTQYSRA VFDGCHYYYL  
 651 PWGNVKKPVV LSSYWEDISH RIETQNQLLG IADRLEAGLE QVHLALKAQC  
 701 STEDVDSLVA QLTDELIAGC SQPLGDIHET PSATHLDOYL YQLRTHHLSQ  
 751 ITEAALALKL GHSELPAALE QAEDWLLRLR ALAEPPQNSL PDIVIWMLQG  
 801 DKRVAYQRPV AHQVLFSTRG ANYCGKNCCK LQITFLKYPM EKVPGARMPV  
 851 QIRVKLWFGV SVDEKEFNQF AEGKLSVFAE TYENETKLAL VGNWGTGTLT  
 901 YPKFSDVTGK IKLPKDSFRP SAGWTWAGDW FVCPEKTLH DMDAGHLSFV  
 951 BEVFENQTRL PGGQWIYMSD NYTDVNGEXV LPKDDIECPL GWKWEDEEWS  
 1001 TDLNRAVDEQ GWEYSITIEP ~~ERKPKHWVPA~~ ~~EKMYTYTERR~~ RWVRLRRRDL  
 1051 ~~SOMEALKRHR~~ QAEAECEGWE YASLFGWKFH LEYRKTDAFR ~~RRWRRRMEP~~  
 1101 LEKTGPAAVF ALEGALGGVM DDKSEDSMSV STLSTGVNRP TISCIFDYGN  
 1151 RYHLRCYMYQ ~~ARDLAAMDKD~~ ~~SFSDPVAIVS~~ ~~FLHOSOXTVV~~ ~~VYNTLNPTWD~~  
 1201 ~~QTLIFYEIEI~~ ~~EGERATVAEQ~~ ~~PPSIVVELVD~~ ~~HDTYGADEEM~~ GRCICQPSLE  
 1251 RMPRLAWFPL TRGSQPSGEL LASFELIQRE KPAIHHPGF EVQETSRILD  
 1301 ESEDTDLPYP PPQREANIYM VPONIKPALQ RTAIEILAWG LRNMKSYQLA  
 1351 NISSPSLVVE CGGQTVQSCV IRNLRKNPNF DICTLFMEVM LPREELYCPS  
 1401 ITVKVIDNRQ FGRRPVVGQC TIRSLESFLC DPYSAESPSP QGGPDDVSL  
 1451 SPGEDVLIDI DDKEPLIPIQ EEEFIDWWSK FFASIGEREK CGSYLEKDFD  
 1501 TLKVYDTQLE NVEAFEGLSF FCNTFKLYRG KTOEETEDPS VIGEFKGLFX  
 1551 IYPLPEDPAI PMPPROFHQL AAQGPQECV RIVIVPAGL ~~QPKDPNGKCD~~  
 1601 ~~PYKHSISGK~~ ~~SVSDODNYIP~~ ~~CTLEPVFGKM~~ ~~FEETCTLPLE~~ ~~KDLKITLYDV~~  
 1651 ~~DLISKDEKIG~~ ETVVDLENRL LSKFGARCGP PQTYCVSGPN QWRDQLRPSQ  
 1701 LLHLFCQQRH VKAPVYRTDR VMFQDKEYSI EEIEAGRI PN PHLGPVEERL  
 1751 ALHVLQOQGL VPEHVESRPL YSPLQPDIEQ GKLMWVDFL PKALGRPGPP  
 1801 FNITERRARR EFLRCIIWNT RDVILDDLSL TGEKMSDIYV KGWMIGFEEH  
 1851 KQKTDVHYRS LGGEGNFNWR FIFPFDYLP EQVCTIAKKD AFWRLDKTES  
 1901 KIPARVVFOI WDNDKFSFDD FLGSLQLDLN RMPKPAKTAK KCSLDQLDDA  
 1951 FHPEWVSLF EQKTVKGWVP CVAEEGEKKI LAGKLEMTLE IVAESEHEER  
 2001 PAGQGRDEPN MNPKLEDPFR PDTSTFLWFTS PYKTMKFIW RFRFWAILF  
 2051 IILFILLLFL AIFIYAFPNY AAMKLEPES  
 (SEQ ID NO:2)

FIG. 2

SUBSTITUTE SHEET (RULE 26)

4/10

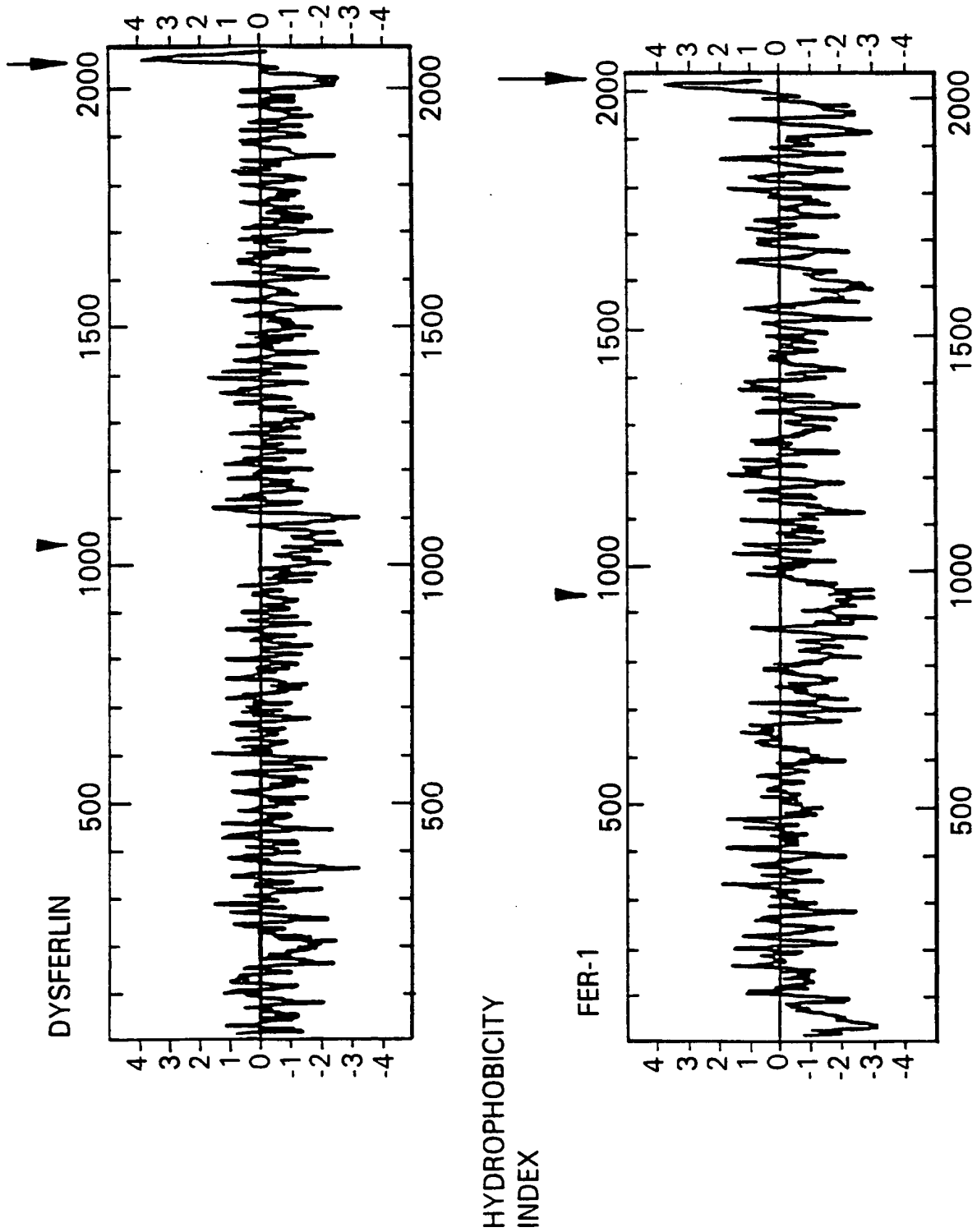
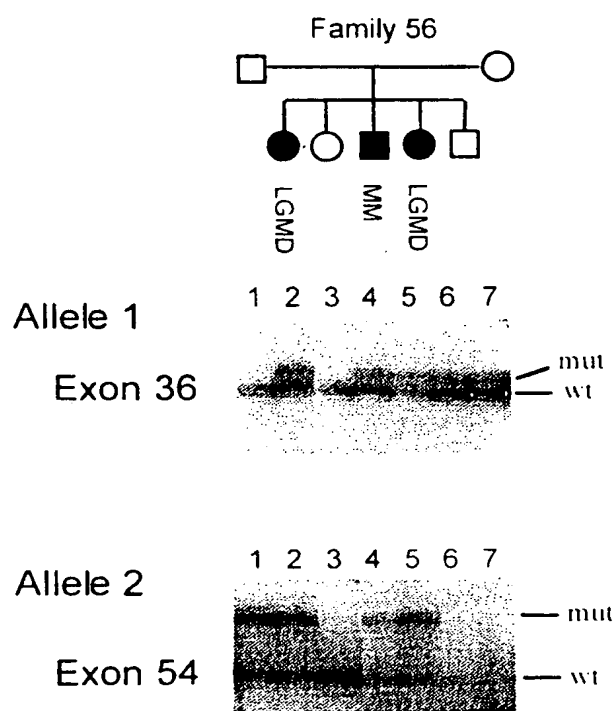


FIG. 3

5/10

**FIG. 4**

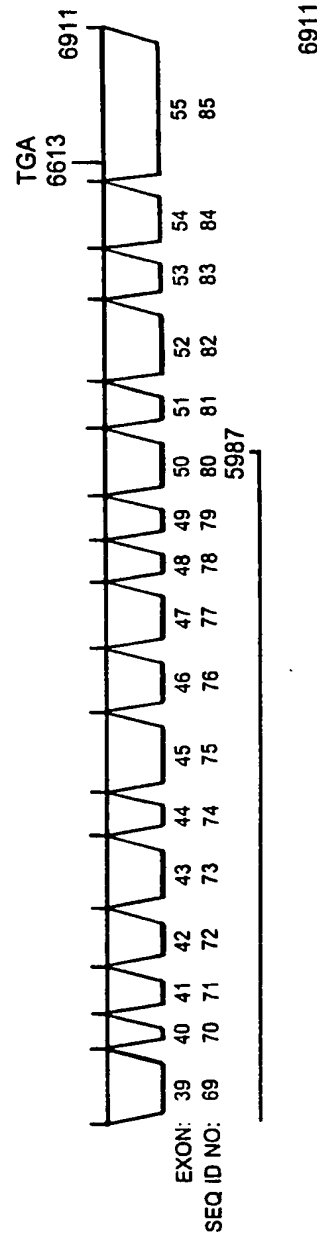
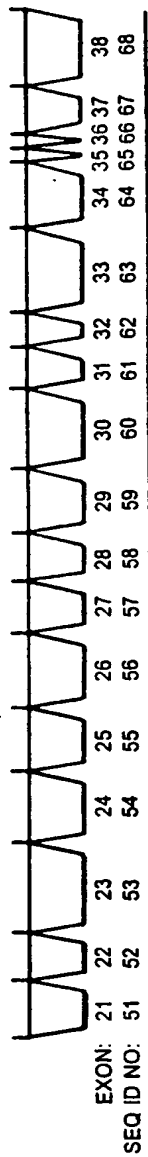
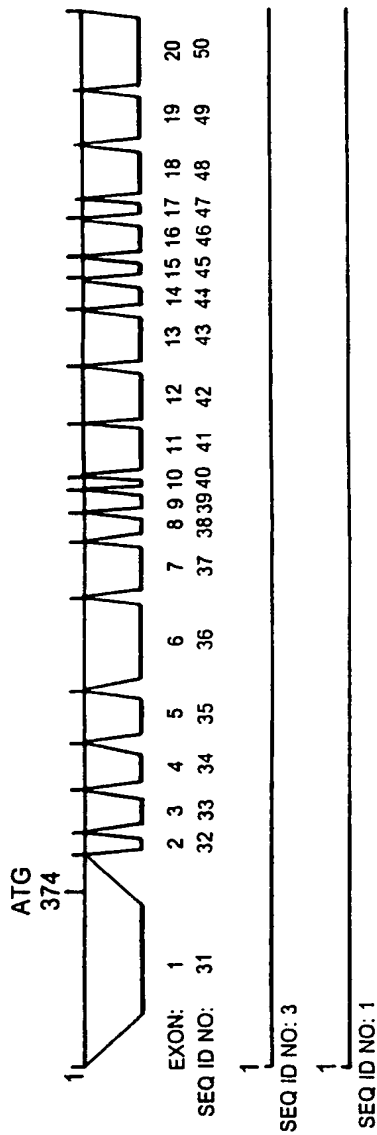


FIG. 5

|         |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1/1     | TCC | TGG | TTC | AAG | CGA | TTC | TCT | TCT | GGC | CTC | AGC | CTC | CCG | AGT | AGC | TGG | GAT | TAC | AGG | CAT | GCT | CCA | CCG | AGC | CCG | GGT | AAT | TTT | GTA | TTT | TTA |
| 31/11   | S   | W   | F   | K   | R   | F   | S   | G   | L   | S   | L   | P   | S   | S   | S   | W   | D   | Y   | R   | H   | A   | P   | P   | S   | P   | G   | N   | F   | V   | F   | L   |
| 91/31   | ATA | GAG | ACG | GGG | TTT | TGC | CAT | GTT | GGT | CAG | GCT | GGT | CTC | GAA | CTC | CTG | ACC | TCA | GGT | GAT | CTG | CCC | ACC | TTG | GCC | TCC | CAA | CGT | GCT | GAG |     |
| 121/41  | I   | E   | T   | G   | F   | C   | H   | V   | G   | Q   | A   | G   | L   | E   | L   | L   | T   | S   | G   | D   | L   | P   | T   | L   | A   | S   | Q   | R   | A   | E   |     |
| 211/71  | ATT | ACA | GGC | ATG | AGT | CAC | TGT | GCC | CGG | CAG | AGA | TGG | TCT | AAT | TCA | TAT | GAA | AGA | ACT | CTG | AGA | AAA | GTA | GAA | AGT | GAT | TTT | CTA | AAA | TAA |     |
| 181/61  | I   | T   | G   | M   | S   | H   | C   | A   | R   | Q   | R   | W   | S   | N   | S   | I   | E   | R   | T   | L   | K   | K   | V   | E   | S   | D   | F   | L   | K   | *   |     |
| 271/91  | GGT | ACA | AAT | AAT | TAA | TGT | AAG | CAT | AAT | CAC | CTA | ACC | TTG | TGG | AAT | TTT | TTT | TTT | TTG | AGA | AGC | AAA | TTG | CAA | ATT | TGT | GAT | AGA | TCT | AAA |     |
| 361/121 | G   | T   | N   | N   | *   | C   | K   | H   | N   | H   | L   | T   | L   | W   | N   | F   | F   | F   | L   | R   | S   | K   | L   | Q   | I   | C   | D   | R   | S   | K   |     |
| 451/151 | TGT | GTG | AAC | CCA | GGA | GGC | AGA | GGT | AGA | GAT | CCA | GAT | P   | G   | E   | G   | V   | M   | D   | D   | K   | S   | E   | D   | S   | M   | S   | V   | S   | T   | L   |
| 541/181 | AGC | tcc | ggg | gtg | aac | aga | ccc | acg | att | tcc | tgc | ata | tcc | gac | tat | ggg | aac | cgc | tac | cat | cta | cgc | tgc | tac | atg | tac | cag | gcc | cgg | gac |     |
| 631/211 | gct | gct | gct | gct | gct | gct | gct | gct | gct | gct | gct | gct | gct | gct | gct | gct | gct | gct | gct | gct | gct | gct | gct | gct | gct | gct | gct | gct | gct | gct |     |
| 721/241 | acc | ctt | aac | ccc | acc | tgg | gac | cag | acg | ctc | atc | ttc | tac | gag | atc | gag | atc | ttt | ggc | gag | ccc | gac | aca | ggt | gct | gag | caa | ccg | ccc | agc |     |

**FIG. 6A**

|   |   |   |
|---|---|---|
| 811/271   | 841/281   | 871/291   |
| att gtc gtc gag gtc tac gac cat gac act tat ggt gca gac gag ttt atg ggt cgc tgc atc tgt caa ccg agt ctg gaa cgc atg cca | att gtc gtc gag gtc tac gac cat gac act tat ggt gca gac gag ttt atg ggt cgc tgc atc tgt caa ccg agt ctg gaa cgc atg cca | att gtc gtc gag gtc tac gac cat gac act tat ggt gca gac gag ttt atg ggt cgc tgc atc tgt caa ccg agt ctg gaa cgc atg cca |
| I V V E L Y D H D T Y C A D E F M G R C I C Q P S L E R M P   | I V V E L Y D H D T Y C A D E F M G R C I C Q P S L E R M P   | I V V E L Y D H D T Y C A D E F M G R C I C Q P S L E R M P   |
| 901/301   | 931/311   | 961/321   |
| cgg ctg gcc tgg ttc cca ctg acg agg ggc agc cag ccg tcg ggg gag ctg ctg gcc tct ttt gag ctc atc cag aga gag aag ccg gcc | cgg ctg gcc tgg ttc cca ctg acg agg ggc agc cag ccg tcg ggg gag ctg ctg gcc tct ttt gag ctc atc cag aga gag aag ccg gcc | cgg ctg gcc tgg ttc cca ctg acg agg ggc agc cag ccg tcg ggg gag ctg ctg gcc tct ttt gag ctc atc cag aga gag aag ccg gcc |
| R L A W F P L T R G S Q P S G E L L A S F E L I Q R E K P A   | R L A W F P L T R G S Q P S G E L L A S F E L I Q R E K P A   | R L A W F P L T R G S Q P S G E L L A S F E L I Q R E K P A   |
| 991/331   | 1021/341  | 1051/351  |
| atc cac cat att cct ggt ttt gag gtg cag gca tca agg atc ctg gat gag tct gag gac aca gac ctg ccc tac cca cca ccc cag     | atc cac cat att cct ggt ttt gag gtg cag gca tca agg atc ctg gat gag tct gag gac aca gac ctg ccc tac cca cca ccc cag     | atc cac cat att cct ggt ttt gag gtg cag gca tca agg atc ctg gat gag tct gag gac aca gac ctg ccc tac cca cca ccc cag     |
| I H H I P G F E V Q E T S R I L D E S E D T D L P Y P P P Q   | I H H I P G F E V Q E T S R I L D E S E D T D L P Y P P P Q   | I H H I P G F E V Q E T S R I L D E S E D T D L P Y P P P Q   |
| 1081/361  | 1111/371  | 1141/381  |
| agg gag gcc aac atc tac atg gtt cct cag aac atc aag cca gcg ctc cag agt acc gcc atc gac atc ctg gca tgg ggc ctg cgg aac | agg gag gcc aac atc tac atg gtt cct cag aac atc aag cca gcg ctc cag agt acc gcc atc gac atc ctg gca tgg ggc ctg cgg aac | agg gag gcc aac atc tac atg gtt cct cag aac atc aag cca gcg ctc cag agt acc gcc atc gac atc ctg gca tgg ggc ctg cgg aac |
| E A N I Y M V P Q N I K P A L Q R T A I E I L A W G L R N   | E A N I Y M V P Q N I K P A L Q R T A I E I L A W G L R N   | E A N I Y M V P Q N I K P A L Q R T A I E I L A W G L R N   |
| 1171/391  | 1201/401  | 1231/411  |
| atg aag agt tac cag ctg gcc aac atc tcc tcc ccc agc ctc gtc gta gag tgt ggg ggc cag acg gtc cag tcc tgt gtc atc agg aac | atg aag agt tac cag ctg gcc aac atc tcc tcc ccc agc ctc gtc gta gag tgt ggg ggc cag acg gtc cag tcc tgt gtc atc agg aac | atg aag agt tac cag ctg gcc aac atc tcc tcc ccc agc ctc gtc gta gag tgt ggg ggc cag acg gtc cag tcc tgt gtc atc agg aac |
| M K S Y Q L A N I S S P S L V V E C G G Q T V Q S C V I R N   | M K S Y Q L A N I S S P S L V V E C G G Q T V Q S C V I R N   | M K S Y Q L A N I S S P S L V V E C G G Q T V Q S C V I R N   |
| 1261/421  | 1291/431  | 1321/441  |
| ctc cgg aag aac ccc aac ttt gac atc tgc acc ctc ttc atg gaa gtg atg ctg ccc agg gag gag ctc tac tgc ccc ccc atc acc gtc | ctc cgg aag aac ccc aac ttt gac atc tgc acc ctc ttc atg gaa gtg atg ctg ccc agg gag gag ctc tac tgc ccc ccc atc acc gtc | ctc cgg aag aac ccc aac ttt gac atc tgc acc ctc ttc atg gaa gtg atg ctg ccc agg gag gag ctc tac tgc ccc ccc atc acc gtc |
| L R K N P N F D I C T L F M E V M L P R E E L Y C P I T V   | L R K N P N F D I C T L F M E V M L P R E E L Y C P I T V   | L R K N P N F D I C T L F M E V M L P R E E L Y C P I T V   |
| 1351/451  | 1381/461  | 1411/471  |
| agc gtc atc gat aac cgc cag ttt ggc cgc cgg cct gtc gtc ggc cag tgt acc atc cgc tcc ctg gag agc ttc ctg tgt gac ccc tac | agc gtc atc gat aac cgc cag ttt ggc cgc cgg cct gtc gtc ggc cag tgt acc atc cgc tcc ctg gag agc ttc ctg tgt gac ccc tac | agc gtc atc gat aac cgc cag ttt ggc cgc cgg cct gtc gtc ggc cag tgt acc atc cgc tcc ctg gag agc ttc ctg tgt gac ccc tac |
| K V I D N R Q F G R P V V G Q C T I R S L E S F L C D P Y   | K V I D N R Q F G R P V V G Q C T I R S L E S F L C D P Y   | K V I D N R Q F G R P V V G Q C T I R S L E S F L C D P Y   |
| 1441/481  | 1471/491  | 1501/501  |
| tcg cgc gag agt cca tcc cca cag ggt ggc cca gac gat gtc agc cta ctc agt cct ggg gaa gac gtc ctc atc gac att gat gac aag | tcg cgc gag agt cca tcc cca cag ggt ggc cca gac gat gtc agc cta ctc agt cct ggg gaa gac gtc ctc atc gac att gat gac aag | tcg cgc gag agt cca tcc cca cag ggt ggc cca gac gat gtc agc cta ctc agt cct ggg gaa gac gtc ctc atc gac att gat gac aag |
| S A E S P S P Q G P D D V S L L S P G E D V L I D I D D K   | S A E S P S P Q G P D D V S L L S P G E D V L I D I D I D D K   | S A E S P S P Q G P D D V S L L S P G E D V L I D I D I D D K   |
| 1531/511  | 1561/521  | 1591/531  |
| gag ccc ctc atc ccc atc cag gag gaa gag ttc atc gat tgg tgg agc aaa ttc ttt gcc tcc ata ggg gag agg gaa aag tgc ggc tcc | gag ccc ctc atc ccc atc cag gag gaa gag ttc atc gat tgg tgg agc aaa ttc ttt gcc tcc ata ggg gag agg gaa aag tgc ggc tcc | gag ccc ctc atc ccc atc cag gag gaa gag ttc atc gat tgg tgg agc aaa ttc ttt gcc tcc ata ggg gag agg gaa aag tgc ggc tcc |
| E P L I P I Q E E F I D W S K F F A S I G E R E K C G S   | E P L I P I Q E E F I D W S K F F A S I G E R E K C G S   | E P L I P I Q E E F I D W S K F F A S I G E R E K C G S   |
| 1621/541  | 1651/551  | 1681/561  |
| tac ctg gag aag gat ttt gac acc ctg aag gtc tat gac aca cag ctg gag aat gtc gag gcc ttt gag ggc ctg tct gac ttt tgt aac | tac ctg gag aag gat ttt gac acc ctg aag gtc tat gac aca cag ctg gag aat gtc gag gcc ttt gag ggc ctg tct gac ttt tgt aac | tac ctg gag aag gat ttt gac acc ctg aag gtc tat gac aca cag ctg gag aat gtc gag gcc ttt gag ggc ctg tct gac ttt tgt aac |
| Y L E K D F D T L K V Y D T Q L E N V E A F E G L S D F C N   | Y L E K D F D T L K V Y D T Q L E N V E A F E G L S D F C N   | Y L E K D F D T L K V Y D T Q L E N V E A F E G L S D F C N   |

**FIG. 6B**



9/10

1711/571 1741/581 1771/591  
 acc ttc aag ctg tac cgg ggc aag acg cag gag agc aca gaa gat cca tct gtg att ggt gaa ttt aag ggc ctc ttc aaa att tat ccc  
 T F K L Y R G K T Q E E T E D P S V I G E F K G L F K I Y P  
 1801/601 1831/611 1861/621  
 ctc cca gaa gac cca gcc atc ccc atg ccc cca aga cag ttc cac cag ctg gcc gcc cag gga ccc cag gag tgc ttg gtc cgt atc tac  
 L P E D P A I P M P P R Q F H Q L A A Q G P Q E C L V R I Y  
 1891/631 1921/641 1951/651  
 att gtc cga gca ttt ggc ctg cag ccc aag gac ccc aat gga aag tgt gat cct tac atc aag atc tcc ata ggg aag aaa tca gtg agt  
 I V R A F G L Q P K D P N G K C D P Y I K I S I G K K S V S  
 1981/661 2011/671 2041/681  
 gac cag gat aac tac atc ccc tgc acg ctg gag ccc gta ttt gga aag atg ttc gag ctg acc tgc att ctg cct ctg gag aag gac cta  
 D Q D N Y I P C T L E P V F G K M F E L T C T L P L E K D L  
 2071/691 2101/701 2131/711  
 aag atc act ctc tat gac tat gac ctc ctc tcc aag gac gaa aag atc ggt gag acg gtc gtc gac ctg gag aac agg ctg ctg tcc aag  
 K I T L Y D Y D L L S K D E K I G E T V V D L E N R L L S K  
 2161/721 2191/731 2221/741  
 ttt ggg gct cgc tgt gga ctc cca cag acc tac tgt gtc tct gga ccg aac cag tgg cgg gac cag ctc cgc ccc tcc cag ctc ctc cac  
 F G A R C G L P Q T Y C V S G P N Q W R D Q L R P S Q L L H  
 2251/751 2281/761 2311/771  
 ctc ttc tgc cag cag cat aga gtc aag gca cct gtg tac cgg aca gac cgt gta atg ttt cag gat aaa gaa tat tcc att gaa gag ata  
 L F C Q Q H R V K A P V Y R T D R V M F Q D K E Y S I E I  
 2341/781 2371/791 2401/801  
 gag gct ggc agg atc cca aac cca cag ctc ggc cca ctg gag gag cgt ctg gct ctg cat gtg ctt cag cag cag ggc ctg gtc ccg gag  
 E A G R I P N P H L G P V E E R L A L H V L Q Q Q G L V P E  
 2431/811 2461/821 2491/831  
 cac gtg gag tca cgg ccc ctc tac agc ccc ctg cag cca gac atc gag cag ggg aag ctg ctg gtc gag cta ttt ccg aag gcc  
 H V E S R P L Y S P L Q P D I E Q G K L Q M W V D L F P K A  
 2521/841 2551/851 2581/861  
 ctg ggg cgg cct gga cct ctc aac atc acc cca cgg aga gcc aga agg ttt ttc ctg cgt tgt att atc tgg aat acc aga gat gtg  
 L G R P G P P F N I T P R R A R R F F L R C I I W N T R D V  
 2611/871 2641/881 2671/891  
 att ctg gat gac ctg agc ctc acg ggg gag aag atg agc gac att tat gtg aaa ggt tgg atg att ggc ttt gaa gaa cac aag caa aag  
 I L D D L S L T G E K M S D I Y V K G W M I G F E H K Q K

FIG. 6C

10/10

2701/301      2731/911      2761/921  
 acc gac gcg cat tat cgt tcc ctg gga ggt gaa ggc aac ttc aac tgg agg ttc att ttc ccc ttc gac tac ctg cca gct gag caa gtc  
 T D V H Y R S L G G E G N F N W R F I F P F D Y L P A E Q V  
 2791/931      2821/941      2851/951  
 cgt acc att gcc aag aag gat gcc ttc tgg agg ctg gac aay act gag agc aaa atc cca gca cga ctg gtg ttc cag atc tgg gac aat  
 C T I A K K D A F W R L D K T E S K I P A R V V F Q I W D N  
 2881/961      2911/971      2941/981  
 gac aag ttc tcc ttt gat gat ttt ctg ggc tcc ctg cag ctg gat ctg aac cgc atg ccc aag cca gcc aag aca gcc aag aag tgc tcc  
 D K F S F D D F L G S L Q L D L N R M P K P A K T A K C S  
 2971/991      3001/1001      3031/1011  
 ttg gac cag ctg gat gat gct ttc cac cca gaa tgg ttt gtg tcc ctt ttt gag cag aaa aca gtg aag ggc tgg tgg ccc tgc gta gca  
 L D Q L D D A F H P E W F V S L F E Q K T V K G W P C V A  
 3061/1021      3091/1031      3121/1041  
 gaa gag ggt gag aag aaa ata ctg gcg ggc aag ctg gaa atg acc ttg gag att gta gca gag agt gag cat gag gag cgg cct gct ggc  
 E E G E K K I L A G K L E M T L E I V A E S E H E R P A G  
 3151/1051      3181/1061      3211/1071  
 cag ggc cgg gat gag ccc aac atg aac cct aag ctt gag gac cca aag cgc ccc gag acc tcc ttc ctg tgg ttt acc tcc cca tac aag  
 Q G R D E P N M N P K L E D P R R P D T S F L W F T S P Y K  
 3241/1081      3271/1091      3301/1101  
 acc atg aag ttc atc ctg tgg cgg cgt ttc cgg tgg gcc atc atc ttc ttc atc atc ctg ctg ctg ctg ttc ctg ggc atc ttc  
 T M K F I L W R R F R W A I I L F I I L L L L F L A I F  
 3331/1111      3361/1121      3391/1131  
 acc tac gcc ttc ccg aac tat gct gcc aac atg aag ctg gtg aag ccc ttc agc tga gga ctc tcc tgc cct gta gaa ggg gcc gtg ggg tcc  
 I Y A F P N Y A A M K L V K P F S G L S C P V E G A V G S  
 3421/1141      3451/1151      3481/1161  
 cct cca gca tgg gac tgg cct gcc tcc tcc gcc cag ctg ggc gag ctg ctc cag acc tcc tag gcc tga ttg tcc tgc cag ggt ggg cag  
 P P A W D W P A S S A Q L G E L L Q T S A L S C Q G G Q  
 3511/1171      3541/1181      3571/1191  
 acc gac aga tgg acc ggc cca cac tcc cag agt tgc taa cat gga gct ctg aga tca ccc cac ttc cat cat ttc ctt ctc ccc caa ccc  
 T D R W T G P H S Q S C H G A L R S P H F H F L L P Q P  
 3601/1201      3631/1211      3661/1221  
 aac gct ttt ttg gat cag ctg aga cat att tca gta taa aac agt tgg aac cac aaa aaa aaa aa  
 N A F L D Q L R H I S V N S W N H K K K K K

(SEQ ID NO:232)  
(SEQ ID NO:233)

FIG. 6D

1/68

## SEQUENCE LISTING

&lt;110&gt; The General Hospital Corporation

&lt;120&gt; DYSFERLIN, A GENE MUTATED IN DISTAL MYOPATHY AND LIMB GIRDLE MUSCULAR DYSTROPHY

&lt;130&gt; 00786/399W02

&lt;150&gt; US 60/097,927

&lt;151&gt; 1998-08-25

&lt;160&gt; 233

&lt;170&gt; FastSEQ for Windows Version 3.0

&lt;210&gt; 1

&lt;211&gt; 6911

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

&lt;222&gt; (374)...(6613)

```

<400> 1
tcgaccgccc agccaggtgc aaaatgccgt gtcattggga gactccgcag ccggagcatt      60
agattacagc tcgacggagc tcgggaaggg cggcgggggg ggaagatgag cagaagcccc      120
tgttctcgga acgccggtcg acaagcgggg tgagcgcagg cggggcgggg acccagccta      180
gcccaactgga gcagccgggg gtggcccgtt cccctttaag agcaactgct ctaagccagg      240
agccagagat tcgagccggc ctgcccagc cagccctctc cagcgagggg acccacaagc      300
ggcgccctcg ccctcccagc ctttccgagc cctctttgcg ccctgggcgc acggggccct      360
acacgcgccca agc atg ctg agg gtc ttc atc ctc tat gcc gag aac gtc      409
                Met Leu Arg Val Phe Ile Leu Tyr Ala Glu Asn Val
                  1                5                10

cac aca ccc gac acc gac atc agc gat gcc tac tgc tcc gcg gtg ttt      457
His Thr Pro Asp Thr Asp Ile Ser Asp Ala Tyr Cys Ser Ala Val Phe
                15                20                25

gca ggg gtg aag aag aga acc aaa gtc atc aag aac agc gtg aac cct      505
Ala Gly Val Lys Lys Arg Thr Lys Val Ile Lys Asn Ser Val Asn Pro
                30                35                40

gta tgg aat gag gga ttt gaa tgg gac ctc aag ggc atc ccc ctg gac      553
Val Trp Asn Glu Gly Phe Glu Trp Asp Leu Lys Gly Ile Pro Leu Asp
                45                50                55

cag ggc tct gag ctt cat gtg gtg gtc aaa gac cat gag acg atg ggg      601
Gln Gly Ser Glu Leu His Val Val Val Lys Asp His Glu Thr Met Gly
                65                70                75

agg aac agg ttc ctg ggg gaa gcc aag gtc cca ctc cga gag gtc ctc      649
Arg Asn Arg Phe Leu Gly Glu Ala Lys Val Pro Leu Arg Glu Val Leu
                80                85                90

gcc acc cct agt ctg tcc gcc agc ttc aat gcc ccc ctg ctg gac acc      697
Ala Thr Pro Ser Leu Ser Ala Ser Phe Asn Ala Pro Leu Leu Asp Thr
                95                100                105

aag aag cag ccc aca ggg gcc tcg ctg gtc ctg cag gtg tcc tac aca      745
Lys Lys Gln Pro Thr Gly Ala Ser Leu Val Leu Gln Val Ser Tyr Thr
                110                115                120

```

2/68

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| ccg | ctg | cct | gga | gct | gtg | ccc | ctg | ttc | ccg | ccc | cct | act | cct | ctg | gag | 793  |
| Pro | Leu | Pro | Gly | Ala | Val | Pro | Leu | Phe | Pro | Pro | Pro | Thr | Pro | Leu | Glu |      |
| 125 |     |     |     |     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |      |
| ccc | tcc | ccg | act | ctg | cct | gac | ctg | gat | gta | gtg | gca | gac | aca | gga | gga | 841  |
| Pro | Ser | Pro | Thr | Leu | Pro | Asp | Leu | Asp | Val | Val | Ala | Asp | Thr | Gly | Gly |      |
|     |     |     |     | 145 |     |     |     |     | 150 |     |     |     |     | 155 |     |      |
| gag | gaa | gac | aca | gag | gac | cag | gga | ctc | act | gga | gat | gag | gcg | gag | cca | 889  |
| Glu | Glu | Asp | Thr | Glu | Asp | Gln | Gly | Leu | Thr | Gly | Asp | Glu | Ala | Glu | Pro |      |
|     |     |     | 160 |     |     |     |     | 165 |     |     |     |     | 170 |     |     |      |
| ttc | ctg | gat | caa | agc | gga | ggc | ccg | ggg | gct | ccc | acc | acc | cca | agg | aaa | 937  |
| Phe | Leu | Asp | Gln | Ser | Gly | Gly | Pro | Gly | Ala | Pro | Thr | Thr | Pro | Arg | Lys |      |
|     |     | 175 |     |     |     |     | 180 |     |     |     |     | 185 |     |     |     |      |
| cta | cct | tca | cgt | cct | ccg | ccc | cac | tac | ccc | ggg | atc | aaa | aga | aag | cga | 985  |
| Leu | Pro | Ser | Arg | Pro | Pro | Pro | His | Tyr | Pro | Gly | Ile | Lys | Arg | Lys | Arg |      |
|     | 190 |     |     |     |     | 195 |     |     |     |     | 200 |     |     |     |     |      |
| agt | gcg | cct | aca | tct | aga | aag | ctg | ctg | tca | gac | aaa | ccg | cag | gat | ttc | 1033 |
| Ser | Ala | Pro | Thr | Ser | Arg | Lys | Leu | Leu | Ser | Asp | Lys | Pro | Gln | Asp | Phe |      |
| 205 |     |     |     |     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |      |
| cag | atc | agg | gtc | cag | gtg | atc | gag | ggg | cgc | cag | ctg | ccg | ggg | gtg | aac | 1081 |
| Gln | Ile | Arg | Val | Gln | Val | Ile | Glu | Gly | Arg | Gln | Leu | Pro | Gly | Val | Asn |      |
|     |     |     |     | 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |      |
| atc | aag | cct | gtg | gtc | aag | gtt | acc | gct | gca | ggg | cag | acc | aag | cgg | acg | 1129 |
| Ile | Lys | Pro | Val | Val | Lys | Val | Thr | Ala | Ala | Gly | Gln | Thr | Lys | Arg | Thr |      |
|     |     |     | 240 |     |     |     | 245 |     |     |     |     |     | 250 |     |     |      |
| cgg | atc | cac | aag | gga | aac | agc | cca | ctc | ttc | aat | gag | act | ctt | ttc | ttc | 1177 |
| Arg | Ile | His | Lys | Gly | Asn | Ser | Pro | Leu | Phe | Asn | Glu | Thr | Leu | Phe | Phe |      |
|     |     | 255 |     |     |     |     | 260 |     |     |     |     | 265 |     |     |     |      |
| aac | ttg | ttt | gac | tct | cct | ggg | gag | ctg | ttt | gat | gag | ccc | atc | ttt | atc | 1225 |
| Asn | Leu | Phe | Asp | Ser | Pro | Gly | Glu | Leu | Phe | Asp | Glu | Pro | Ile | Phe | Ile |      |
|     | 270 |     |     |     |     | 275 |     |     |     |     | 280 |     |     |     |     |      |
| acg | gtg | gta | gac | tct | cgt | tct | ctc | agg | aca | gat | gct | ctc | ctc | ggg | gag | 1273 |
| Thr | Val | Val | Asp | Ser | Arg | Ser | Leu | Arg | Thr | Asp | Ala | Leu | Leu | Gly | Glu |      |
| 285 |     |     |     |     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |      |
| ttc | cgg | atg | gac | gtg | ggc | acc | att | tac | aga | gag | ccc | cgg | cac | gcc | tat | 1321 |
| Phe | Arg | Met | Asp | Val | Gly | Thr | Ile | Tyr | Arg | Glu | Pro | Arg | His | Ala | Tyr |      |
|     |     |     |     | 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |      |
| ctc | agg | aag | tgg | ctg | ctg | ctc | tca | gac | cct | gat | gac | ttc | tct | gct | ggg | 1369 |
| Leu | Arg | Lys | Trp | Leu | Leu | Leu | Ser | Asp | Pro | Asp | Asp | Phe | Ser | Ala | Gly |      |
|     |     |     | 320 |     |     |     |     | 325 |     |     |     |     | 330 |     |     |      |
| gcc | aga | ggc | tac | ctg | aaa | aca | agc | ctt | tgt | gtg | ctg | ggg | cct | ggg | gac | 1417 |
| Ala | Arg | Gly | Tyr | Leu | Lys | Thr | Ser | Leu | Cys | Val | Leu | Gly | Pro | Gly | Asp |      |
|     |     | 335 |     |     |     |     | 340 |     |     |     |     | 345 |     |     |     |      |
| gaa | gcg | cct | ctg | gag | aga | aaa | gac | ccc | tct | gaa | gac | aag | gag | gac | att | 1465 |
| Glu | Ala | Pro | Leu | Glu | Arg | Lys | Asp | Pro | Ser | Glu | Asp | Lys | Glu | Asp | Ile |      |
|     | 350 |     |     |     |     | 355 |     |     |     |     | 360 |     |     |     |     |      |
| gaa | agc | aac | ctg | ctc | cgg | ccc | aca | ggc | gta | gcc | ctg | cga | gga | gcc | cac | 1513 |
| Glu | Ser | Asn | Leu | Leu | Arg | Pro | Thr | Gly | Val | Ala | Leu | Arg | Gly | Ala | His |      |
| 365 |     |     |     |     | 370 |     |     |     |     | 375 |     |     |     |     | 380 |      |
| ttc | tgc | ctg | aag | gtc | ttc | cgg | gcc | gag | gac | ttg | ccg | cag | atg | gac | gat | 1561 |
| Phe | Cys | Leu | Lys | Val | Phe | Arg | Ala | Glu | Asp | Leu | Pro | Gln | Met | Asp | Asp |      |
|     |     |     |     | 385 |     |     |     |     | 390 |     |     |     |     | 395 |     |      |

3/68

|   |      |
|---|------|
| gcc gtg atg gac aac gtg aaa cag atc ttt ggc ttc gag agt aac aag<br>Ala Val Met Asp Asn Val Lys Gln Ile Phe Gly Phe Glu Ser Asn Lys<br>400 405 410     | 1609 |
| aag aac ttg gtg gac ccc ttt gtg gag gtc agc ttt gcg ggg aaa atg<br>Lys Asn Leu Val Asp Pro Phe Val Glu Val Ser Phe Ala Gly Lys Met<br>415 420 425     | 1657 |
| ctg tgc agc aag atc ttg gag aag acg gcc aac cct cag tgg aac cag<br>Leu Cys Ser Lys Ile Leu Glu Lys Thr Ala Asn Pro Gln Trp Asn Gln<br>430 435 440     | 1705 |
| aac atc aca ctg cct gcc atg ttt ccc tcc atg tgc gaa aaa atg agg<br>Asn Ile Thr Leu Pro Ala Met Phe Pro Ser Met Cys Glu Lys Met Arg<br>445 450 455 460 | 1753 |
| att cgt atc ata gac tgg gac cgc ctg act cac aat gac atc gtg gct<br>Ile Arg Ile Ile Asp Trp Asp Arg Leu Thr His Asn Asp Ile Val Ala<br>465 470 475     | 1801 |
| acc acc tac ctg agt atg tgc aaa atc tct gcc cct gga gga gaa ata<br>Thr Thr Tyr Leu Ser Met Ser Lys Ile Ser Ala Pro Gly Gly Glu Ile<br>480 485 490     | 1849 |
| gaa gag gag cct gca ggt gct gtc aag cct tgc aaa gcc tca gac ttg<br>Glu Glu Glu Pro Ala Gly Ala Val Lys Pro Ser Lys Ala Ser Asp Leu<br>495 500 505     | 1897 |
| gat gac tac ctg ggc ttc ctc ccc act ttt ggg ccc tgc tac atc aac<br>Asp Asp Tyr Leu Gly Phe Leu Pro Thr Phe Gly Pro Cys Tyr Ile Asn<br>510 515 520     | 1945 |
| ctc tat ggc agt ccc aga gag ttc aca ggc ttc cca gac ccc tac aca<br>Leu Tyr Gly Ser Pro Arg Glu Phe Thr Gly Phe Pro Asp Pro Tyr Thr<br>525 530 535 540 | 1993 |
| gag ctc aac aca ggc aag ggg gaa ggt gtg gct tat cgt ggc cgg ctt<br>Glu Leu Asn Thr Gly Lys Gly Glu Gly Val Ala Tyr Arg Gly Arg Leu<br>545 550 555     | 2041 |
| ctg ctc tcc ctg gag acc aag ctg gtg gag cac agt gaa cag aag gtg<br>Leu Leu Ser Leu Glu Thr Lys Leu Val Glu His Ser Glu Gln Lys Val<br>560 565 570     | 2089 |
| gag gac ctt cct gcg gat gac atc ctc cgg gtg gag aag tac ctt agg<br>Glu Asp Leu Pro Ala Asp Asp Ile Leu Arg Val Glu Lys Tyr Leu Arg<br>575 580 585     | 2137 |
| agg cgc aag tac tcc ctg ttt gcg gcc ttc tac tca gcc acc atg ctg<br>Arg Arg Lys Tyr Ser Leu Phe Ala Ala Phe Tyr Ser Ala Thr Met Leu<br>590 595 600     | 2185 |
| cag gat gtg gat gat gcc atc cag ttt gag gtc agc atc ggg aac tac<br>Gln Asp Val Asp Asp Ala Ile Gln Phe Glu Val Ser Ile Gly Asn Tyr<br>605 610 615 620 | 2233 |
| ggg aac aag ttc gac atg acc tgc ctg ccg ctg gcc tcc acc act cag<br>Gly Asn Lys Phe Asp Met Thr Cys Leu Pro Leu Ala Ser Thr Thr Gln<br>625 630 635     | 2281 |
| tac agc cgt gca gtc ttt gac ggg tgc cac tac tac tac cta ccc tgg<br>Tyr Ser Arg Ala Val Phe Asp Gly Cys His Tyr Tyr Tyr Leu Pro Trp<br>640 645 650     | 2329 |

4/68

|   |      |
|---|------|
| ggt aac gtg aaa cct gtg gtg gtg ctg tca tcc tac tgg gag gac atc<br>Gly Asn Val Lys Pro Val Val Val Leu Ser Ser Tyr Trp Glu Asp Ile<br>655 660 665     | 2377 |
| agc cat aga atc gag act cag aac cag ctg ctt ggg att gct gac cgg<br>Ser His Arg Ile Glu Thr Gln Asn Gln Leu Leu Gly Ile Ala Asp Arg<br>670 675 680     | 2425 |
| ctg gaa gct ggc ctg gag cag gtc cac ctg gcc ctg aag gcg cag tgc<br>Leu Glu Ala Gly Leu Glu Gln Val His Leu Ala Leu Lys Ala Gln Cys<br>685 690 695 700 | 2473 |
| tcc acg gag gac gtg gac tcg ctg gtg gct cag ctg acg gat gag ctc<br>Ser Thr Glu Asp Val Asp Ser Leu Val Ala Gln Leu Thr Asp Glu Leu<br>705 710 715     | 2521 |
| atc gca ggc tgc agc cag cct ctg ggt gac atc cat gag aca ccc tct<br>Ile Ala Gly Cys Ser Gln Pro Leu Gly Asp Ile His Glu Thr Pro Ser<br>720 725 730     | 2569 |
| gcc acc cac ctg gac cag tac ctg tac cag ctg cgc acc cat cac ctg<br>Ala Thr His Leu Asp Gln Tyr Leu Tyr Gln Leu Arg Thr His His Leu<br>735 740 745     | 2617 |
| agc caa atc act gag gct gcc ctg gcc ctg aag ctc ggc cac agt gag<br>Ser Gln Ile Thr Glu Ala Ala Leu Ala Leu Lys Leu Gly His Ser Glu<br>750 755 760     | 2665 |
| ctc cct gca gct ctg gag cag gcg gag gac tgg ctc ctg cgt ctg cgt<br>Leu Pro Ala Ala Leu Glu Gln Ala Glu Asp Trp Leu Leu Arg Leu Arg<br>765 770 775 780 | 2713 |
| gcc ctg gca gag gag ccc cag aac agc ctg ccg gac atc gtc atc tgg<br>Ala Leu Ala Glu Glu Pro Gln Asn Ser Leu Pro Asp Ile Val Ile Trp<br>785 790 795     | 2761 |
| atg ctg cag gga gac aag cgt gtg gca tac cag cgg gtg ccc gcc cac<br>Met Leu Gln Gly Asp Lys Arg Val Ala Tyr Gln Arg Val Pro Ala His<br>800 805 810     | 2809 |
| caa gtc ctc ttc tcc cgg cgg ggt gcc aac tac tgt ggc aag aat tgt<br>Gln Val Leu Phe Ser Arg Arg Gly Ala Asn Tyr Cys Gly Lys Asn Cys<br>815 820 825     | 2857 |
| ggg aag cta cag aca atc ttt ctg aaa tat ccg atg gag aag gtg cct<br>Gly Lys Leu Gln Thr Ile Phe Leu Lys Tyr Pro Met Glu Lys Val Pro<br>830 835 840     | 2905 |
| ggc gcc cgg atg cca gtg cag ata cgg gtc aag ctg tgg ttt ggg ctc<br>Gly Ala Arg Met Pro Val Gln Ile Arg Val Lys Leu Trp Phe Gly Leu<br>845 850 855 860 | 2953 |
| tct gtg gat gag aag gag ttc aac cag ttt gct gag ggg aag ctg tct<br>Ser Val Asp Glu Lys Glu Phe Asn Gln Phe Ala Glu Gly Lys Leu Ser<br>865 870 875     | 3001 |
| gtc ttt gct gaa acc tat gag aac gag act aag ttg gcc ctt gtt ggg<br>Val Phe Ala Glu Thr Tyr Glu Asn Glu Thr Lys Leu Ala Leu Val Gly<br>880 885 890     | 3049 |
| aac tgg ggc aca acg ggc ctc acc tac ccc aag ttt tct gac gtc acg<br>Asn Trp Gly Thr Thr Gly Leu Thr Tyr Pro Lys Phe Ser Asp Val Thr<br>895 900 905     | 3097 |
| ggc aag atc aag cta ccc aag gac agc ttc cgc ccc tcg gcc ggc tgg<br>Gly Lys Ile Lys Leu Pro Lys Asp Ser Phe Arg Pro Ser Ala Gly Trp<br>910 915 920     | 3145 |

5/68

|   |      |
|---|------|
| acc tgg gct gga gat tgg ttc gtg tgt ccg gag aag act ctg ctc cat | 3193 |
| Thr Trp Ala Gly Asp Trp Phe Val Cys Pro Glu Lys Thr Leu Leu His |      |
| 925 930 935 940   |      |
| gac atg gac gcc ggt cac ctg agc ttc gtg gaa gag gtg ttt gag aac | 3241 |
| Asp Met Asp Ala Gly His Leu Ser Phe Val Glu Glu Val Phe Glu Asn |      |
| 945 950 955   |      |
| cag acc cgg ctt ccc gga ggc cag tgg atc tac atg agt gac aac tac | 3289 |
| Gln Thr Arg Leu Pro Gly Gly Gln Trp Ile Tyr Met Ser Asp Asn Tyr |      |
| 960 965 970   |      |
| acc gat gtg aac ggg gag aag gtg ctt ccc aag gat gac att gag tgc | 3337 |
| Thr Asp Val Asn Gly Glu Lys Val Leu Pro Lys Asp Asp Ile Glu Cys |      |
| 975 980 985   |      |
| cca ctg ggc tgg aag tgg gaa gat gag gaa tgg tcc aca gac ctc aac | 3385 |
| Pro Leu Gly Trp Lys Trp Glu Asp Glu Glu Trp Ser Thr Asp Leu Asn |      |
| 990 995 1000  |      |
| cgg gct gtc gat gag caa ggc tgg gag tat agc atc acc atc ccc ccg | 3433 |
| Arg Ala Val Asp Glu Gln Gly Trp Glu Tyr Ser Ile Thr Ile Pro Pro |      |
| 1005 1010 1015 1020   |      |
| gag cgg aag ccg aag cac tgg gtc cct gct gag aag atg tac tac aca | 3481 |
| Glu Arg Lys Pro Lys His Trp Val Pro Ala Glu Lys Met Tyr Tyr Thr |      |
| 1025 1030 1035  |      |
| cac cga cgg cgg cgc tgg gtg cgc ctg cgc agg agg gat ctc agc caa | 3529 |
| His Arg Arg Arg Arg Trp Val Arg Leu Arg Arg Arg Asp Leu Ser Gln |      |
| 1040 1045 1050  |      |
| atg gaa gca ctg aaa agg cac agg cag gcg gag gcg gag ggc gag ggc | 3577 |
| Met Glu Ala Leu Lys Arg His Arg Gln Ala Glu Ala Glu Gly Glu Gly |      |
| 1055 1060 1065  |      |
| tgg gag tac gcc tct ctt ttt ggc tgg aag ttc cac ctc gag tac cgc | 3625 |
| Trp Glu Tyr Ala Ser Leu Phe Gly Trp Lys Phe His Leu Glu Tyr Arg |      |
| 1070 1075 1080  |      |
| aag aca gat gcc ttc cgc cgc cgc cgc tgg cgc cgt cgc atg gag cca | 3673 |
| Lys Thr Asp Ala Phe Arg Arg Arg Arg Trp Arg Arg Arg Met Glu Pro |      |
| 1085 1090 1095 1100   |      |
| ctg gag aag acg ggg cct gca gct gtg ttt gcc ctt gag ggg gcc ctg | 3721 |
| Leu Glu Lys Thr Gly Pro Ala Ala Val Phe Ala Leu Glu Gly Ala Leu |      |
| 1105 1110 1115  |      |
| ggc ggc gtg atg gat gac aag agt gaa gat tcc atg tcc gtc tcc acc | 3769 |
| Gly Gly Val Met Asp Asp Lys Ser Glu Asp Ser Met Ser Val Ser Thr |      |
| 1120 1125 1130  |      |
| ttg agc ttc ggt gtg aac aga ccc acg att tcc tgc ata ttc gac tat | 3817 |
| Leu Ser Phe Gly Val Asn Arg Pro Thr Ile Ser Cys Ile Phe Asp Tyr |      |
| 1135 1140 1145  |      |
| ggg aac cgc tac cat cta cgc tgc tac atg tac cag gcc cgg gac ctg | 3865 |
| Gly Asn Arg Tyr His Leu Arg Cys Tyr Met Tyr Gln Ala Arg Asp Leu |      |
| 1150 1155 1160  |      |
| gct gcg atg gac aag gac tct ttt tct gat ccc tat gcc atc gtc tcc | 3913 |
| Ala Ala Met Asp Lys Asp Ser Phe Ser Asp Pro Tyr Ala Ile Val Ser |      |
| 1165 1170 1175 1180   |      |

6/68

|   |      |
|---|------|
| ttc ctg cac cag agc cag aag acg gtg gtg gtg aag aac acc ctt aac | 3961 |
| Phe Leu His Gln Ser Gln Lys Thr Val Val Val Lys Asn Thr Leu Asn |      |
| 1185 1190 1195  |      |
| ccc acc tgg gac cag acg ctc atc ttc tac gag atc gag atc ttt ggc | 4009 |
| Pro Thr Trp Asp Gln Thr Leu Ile Phe Tyr Glu Ile Glu Ile Phe Gly |      |
| 1200 1205 1210  |      |
| gag ccg gcc aca gtt gct gag caa ccg ccc agc att gtg gtg gag ctg | 4057 |
| Glu Pro Ala Thr Val Ala Glu Gln Pro Pro Ser Ile Val Val Glu Leu |      |
| 1215 1220 1225  |      |
| tac gac cat gac act tat ggt gca gac gag ttt atg ggt cgc tgc atc | 4105 |
| Tyr Asp His Asp Thr Tyr Gly Ala Asp Glu Phe Met Gly Arg Cys Ile |      |
| 1230 1235 1240  |      |
| tgt caa ccg agt ctg gaa cgg atg cca cgg ctg gcc tgg ttc cca ctg | 4153 |
| Cys Gln Pro Ser Leu Glu Arg Met Pro Arg Leu Ala Trp Phe Pro Leu |      |
| 1245 1250 1255 1260   |      |
| acg agg ggc agc cag ccg tcg ggg gag ctg ctg gcc tct ttt gag ctc | 4201 |
| Thr Arg Gly Ser Gln Pro Ser Gly Glu Leu Leu Ala Ser Phe Glu Leu |      |
| 1265 1270 1275  |      |
| atc cag aga gag aag ccg gcc atc cac cat att cct ggt ttt gag gtg | 4249 |
| Ile Gln Arg Glu Lys Pro Ala Ile His His Ile Pro Gly Phe Glu Val |      |
| 1280 1285 1290  |      |
| cag gag aca tca agg atc ctg gat gag tct gag gac aca gac ctg ccc | 4297 |
| Gln Glu Thr Ser Arg Ile Leu Asp Glu Ser Glu Asp Thr Asp Leu Pro |      |
| 1295 1300 1305  |      |
| tac cca cca ccc cag agg gag gcc aac atc tac atg gtt cct cag aac | 4345 |
| Tyr Pro Pro Pro Gln Arg Glu Ala Asn Ile Tyr Met Val Pro Gln Asn |      |
| 1310 1315 1320  |      |
| atc aag cca gcg ctc cag cgt acc gcc atc gag atc ctg gca tgg ggc | 4393 |
| Ile Lys Pro Ala Leu Gln Arg Thr Ala Ile Glu Ile Leu Ala Trp Gly |      |
| 1325 1330 1335 1340   |      |
| ctg cgg aac atg aag agt tac cag ctg gcc aac atc tcc tcc ccc agc | 4441 |
| Leu Arg Asn Met Lys Ser Tyr Gln Leu Ala Asn Ile Ser Ser Pro Ser |      |
| 1345 1350 1355  |      |
| ctc gtg gta gag tgt ggg ggc cag acg gtg cag tcc tgt gtc atc agg | 4489 |
| Leu Val Val Glu Cys Gly Gly Gln Thr Val Gln Ser Cys Val Ile Arg |      |
| 1360 1365 1370  |      |
| aac ctc cgg aag aac ccc aac ttt gac atc tgc acc ctc ttc atg gaa | 4537 |
| Asn Leu Arg Lys Asn Pro Asn Phe Asp Ile Cys Thr Leu Phe Met Glu |      |
| 1375 1380 1385  |      |
| gtg atg ctg ccc agg gag gag ctc tac tgc ccc ccc atc acc gtc aag | 4585 |
| Val Met Leu Pro Arg Glu Glu Leu Tyr Cys Pro Pro Ile Thr Val Lys |      |
| 1390 1395 1400  |      |
| gtc atc gat aac cgc cag ttt ggc cgc cgg cct gtg gtg ggc cag tgt | 4633 |
| Val Ile Asp Asn Arg Gln Phe Gly Arg Arg Pro Val Val Gly Gln Cys |      |
| 1405 1410 1415 1420   |      |
| acc atc cgc tcc ctg gag agc ttc ctg tgt gac ccc tac tcg gcg gag | 4681 |
| Thr Ile Arg Ser Leu Glu Ser Phe Leu Cys Asp Pro Tyr Ser Ala Glu |      |
| 1425 1430 1435  |      |
| agt cca tcc cca cag ggt ggc cca gac gat gtg agc cta ctc agt cct | 4729 |
| Ser Pro Ser Pro Gln Gly Gly Pro Asp Asp Val Ser Leu Leu Ser Pro |      |
| 1440 1445 1450  |      |



7/68

|   |      |
|---|------|
| ggg gaa gac gtg ctc atc gac att gat gac aag gag ccc ctc atc ccc<br>Gly Glu Asp Val Leu Ile Asp Ile Asp Asp Lys Glu Pro Leu Ile Pro<br>1455 1460 1465      | 4777 |
| atc cag gag gaa gag ttc atc gat tgg tgg agc aaa ttc ttt gcc tcc<br>Ile Gln Glu Glu Glu Phe Ile Asp Trp Trp Ser Lys Phe Phe Ala Ser<br>1470 1475 1480      | 4825 |
| ata ggg gag agg gaa aag tgc ggc tcc tac ctg gag aag gat ttt gac<br>Ile Gly Glu Arg Glu Lys Cys Gly Ser Tyr Leu Glu Lys Asp Phe Asp<br>1485 1490 1495 1500 | 4873 |
| acc ctg aag gtc tat gac aca cag ctg gag aat gtg gag gcc ttt gag<br>Thr Leu Lys Val Tyr Asp Thr Gln Leu Glu Asn Val Glu Ala Phe Glu<br>1505 1510 1515      | 4921 |
| ggc ctg tct gac ttt tgt aac acc ttc aag ctg tac cgg ggc aag acg<br>Gly Leu Ser Asp Phe Cys Asn Thr Phe Lys Leu Tyr Arg Gly Lys Thr<br>1520 1525 1530      | 4969 |
| cag gag gag aca gaa gat cca tct gtg att ggt gaa ttt aag ggc ctc<br>Gln Glu Glu Thr Glu Asp Pro Ser Val Ile Gly Glu Phe Lys Gly Leu<br>1535 1540 1545      | 5017 |
| ttc aaa att tat ccc ctc cca gaa gac cca gcc atc ccc atg ccc cca<br>Phe Lys Ile Tyr Pro Leu Pro Glu Asp Pro Ala Ile Pro Met Pro Pro<br>1550 1555 1560      | 5065 |
| aga cag ttc cac cag ctg gcc gcc cag gga ccc cag gag tgc ttg gtc<br>Arg Gln Phe His Gln Leu Ala Ala Gln Gly Pro Gln Glu Cys Leu Val<br>1565 1570 1575 1580 | 5113 |
| cgt atc tac att gtc cga gca ttt ggc ctg cag ccc aag gac ccc aat<br>Arg Ile Tyr Ile Val Arg Ala Phe Gly Leu Gln Pro Lys Asp Pro Asn<br>1585 1590 1595      | 5161 |
| gga aag tgt gat cct tac atc aag atc tcc ata ggg aag aaa tca gtg<br>Gly Lys Cys Asp Pro Tyr Ile Lys Ile Ser Ile Gly Lys Lys Ser Val<br>1600 1605 1610      | 5209 |
| agt gac cag gat aac tac atc ccc tgc acg ctg gag ccc gta ttt gga<br>Ser Asp Gln Asp Asn Tyr Ile Pro Cys Thr Leu Glu Pro Val Phe Gly<br>1615 1620 1625      | 5257 |
| aag atg ttc gag ctg acc tgc act ctg cct ctg gag aag gac cta aag<br>Lys Met Phe Glu Leu Thr Cys Thr Leu Pro Leu Glu Lys Asp Leu Lys<br>1630 1635 1640      | 5305 |
| atc act ctc tat gac tat gac ctc ctc tcc aag gac gaa aag atc ggt<br>Ile Thr Leu Tyr Asp Tyr Asp Leu Leu Ser Lys Asp Glu Lys Ile Gly<br>1645 1650 1655 1660 | 5353 |
| gag acg gtc gtc gac ctg gag aac agg ctg ctg tcc aag ttt ggg gct<br>Glu Thr Val Val Asp Leu Glu Asn Arg Leu Leu Ser Lys Phe Gly Ala<br>1665 1670 1675      | 5401 |
| cgc tgt gga ctc cca cag acc tac tgt gtc tct gga ccg aac cag tgg<br>Arg Cys Gly Leu Pro Gln Thr Tyr Cys Val Ser Gly Pro Asn Gln Trp<br>1680 1685 1690      | 5449 |
| cgg gac cag ctc cgc ccc tcc cag ctc ctc cac ctc ttc tgc cag cag<br>Arg Asp Gln Leu Arg Pro Ser Gln Leu Leu His Leu Phe Cys Gln Gln<br>1695 1700 1705      | 5497 |

8/68

|   |      |
|---|------|
| cat aga gtc aag gca cct gtg tac cgg aca gac cgt gta atg ttt cag<br>His Arg Val Lys Ala Pro Val Tyr Arg Thr Asp Arg Val Met Phe Gln<br>1710 1715 1720      | 5545 |
| gat aaa gaa tat tcc att gaa gag ata gag gct ggc agg atc cca aac<br>Asp Lys Glu Tyr Ser Ile Glu Glu Ile Glu Ala Gly Arg Ile Pro Asn<br>1725 1730 1735 1740 | 5593 |
| cca cac ctg ggc cca gtg gag gag cgt ctg gct ctg cat gtg ctt cag<br>Pro His Leu Gly Pro Val Glu Glu Arg Leu Ala Leu His Val Leu Gln<br>1745 1750 1755      | 5641 |
| cag cag ggc ctg gtc ccg gag cac gtg gag tca cgg ccc ctc tac agc<br>Gln Gln Gly Leu Val Pro Glu His Val Glu Ser Arg Pro Leu Tyr Ser<br>1760 1765 1770      | 5689 |
| ccc ctg cag cca gac atc gag cag ggg aag ctg cag atg tgg gtc gac<br>Pro Leu Gln Pro Asp Ile Glu Gln Gly Lys Leu Gln Met Trp Val Asp<br>1775 1780 1785      | 5737 |
| cta ttt ccg aag gcc ctg ggg cgg cct gga cct ccc ttc aac atc acc<br>Leu Phe Pro Lys Ala Leu Gly Arg Pro Gly Pro Pro Phe Asn Ile Thr<br>1790 1795 1800      | 5785 |
| cca cgg aga gcc aga agg ttt ttc ctg cgt tgt att atc tgg aat acc<br>Pro Arg Arg Ala Arg Arg Phe Phe Leu Arg Cys Ile Ile Trp Asn Thr<br>1805 1810 1815 1820 | 5833 |
| aga gat gtg atc ctg gat gac ctg agc ctc acg ggg gag aag atg agc<br>Arg Asp Val Ile Leu Asp Asp Leu Ser Thr Gly Glu Lys Met Ser<br>1825 1830 1835          | 5881 |
| gac att tat gtg aaa ggt tgg atg att ggc ttt gaa gaa cac aag caa<br>Asp Ile Tyr Val Lys Gly Trp Met Ile Gly Phe Glu Glu His Lys Gln<br>1840 1845 1850      | 5929 |
| aag aca gac gtg cat tat cgt tcc ctg gga ggt gaa ggc aac ttc aac<br>Lys Thr Asp Val His Tyr Arg Ser Leu Gly Gly Glu Gly Asn Phe Asn<br>1855 1860 1865      | 5977 |
| tgg agg ttc att ttc ccc ttc gac tac ctg cca gct gag caa gtc tgt<br>Trp Arg Phe Ile Phe Pro Phe Asp Tyr Leu Pro Ala Glu Gln Val Cys<br>1870 1875 1880      | 6025 |
| acc att gcc aag aag gat gcc ttc tgg agg ctg gac aag act gag agc<br>Thr Ile Ala Lys Lys Asp Ala Phe Trp Arg Leu Asp Lys Thr Glu Ser<br>1885 1890 1895 1900 | 6073 |
| aaa atc cca gca cga gtg gtg ttc cag atc tgg gac aat gac aag ttc<br>Lys Ile Pro Ala Arg Val Val Phe Gln Ile Trp Asp Asn Asp Lys Phe<br>1905 1910 1915      | 6121 |
| tcc ttt gat gat ttt ctg ggc tcc ctg cag ctc gat ctc aac cgc atg<br>Ser Phe Asp Asp Phe Leu Gly Ser Leu Gln Leu Asp Leu Asn Arg Met<br>1920 1925 1930      | 6169 |
| ccc aag cca gcc aag aca gcc aag aag tgc tcc ttg gac cag ctg gat<br>Pro Lys Pro Ala Lys Thr Ala Lys Lys Cys Ser Leu Asp Gln Leu Asp<br>1935 1940 1945      | 6217 |
| gat gct ttc cac cca gaa tgg ttt gtg tcc ctt ttt gag cag aaa aca<br>Asp Ala Phe His Pro Glu Trp Phe Val Ser Leu Phe Glu Gln Lys Thr<br>1950 1955 1960      | 6265 |
| gtg aag ggc tgg tgg ccc tgt gta gca gaa gag ggt gag aag aaa ata<br>Val Lys Gly Trp Trp Pro Cys Val Ala Glu Glu Gly Glu Lys Lys Ile<br>1965 1970 1975 1980 | 6313 |

9/68

|   |                                      |
|---|--------------------------------------|
| ctg gcg ggc aag ctg gaa atg acc ttg gag att gta gca gag agt gag<br>Leu Ala Gly Lys Leu Glu Met Thr Leu Glu Ile Val Ala Glu Ser Glu<br>1985 1990 1995  | 6361                                 |
| cat gag gag cgg cct gct ggc cag ggc cgg gat gag ccc aac atg aac<br>His Glu Glu Arg Pro Ala Gly Gln Gly Arg Asp Glu Pro Asn Met Asn<br>2000 2005 2010  | 6409                                 |
| cct aag ctt gag gac cca agg cgc ccc gac acc tcc ttc ctg tgg ttt<br>Pro Lys Leu Glu Asp Pro Arg Arg Pro Asp Thr Ser Phe Leu Trp Phe<br>2015 2020 2025  | 6457                                 |
| acc tcc cca tac aag acc atg aag ttc atc ctg tgg cgg cgt ttc cgg<br>Thr Ser Pro Tyr Lys Thr Met Lys Phe Ile Leu Trp Arg Arg Phe Arg<br>2030 2035 2040  | 6505                                 |
| tgg gcc atc atc ctc ttc atc atc ctc ttc atc ctg ctg ctg ttc ctg<br>Trp Ala Ile Ile Leu Phe Ile Ile Leu Phe Ile Leu Leu Leu Phe Leu<br>2045 2050 2055 2060   | 6553                                 |
| gcc atc ttc atc tac gcc ttc ccg aac tat gct gcc atg aag ctg gtg<br>Ala Ile Phe Ile Tyr Ala Phe Pro Asn Tyr Ala Ala Met Lys Leu Val<br>2065 2070 2075  | 6601                                 |
| aag ccc ttc agc tgaggactct cctgccctgt agaagggggcc gtgggggtccc<br>Lys Pro Phe Ser<br>2080  | 6653                                 |
| ctccagcatg ggactggcct gcctcctccg cccagctcgg cgagctcctc cagacctcct<br>aggcctgatt gtcctgccag ggtgggcaga cagacagatg gaccggccca cactcccaga<br>gttgctaaca tggagctctg agatcacccc acttccatca tttccttctc ccccaaccga<br>acgctttttt ggatcagctc agacatatatt cagtataaaa cagttggaac cacaaaaaaa<br>aaaaaaaaa aaaaaaaa | 6713<br>6773<br>6833<br>6893<br>6911 |

<210> 2  
 <211> 2080  
 <212> PRT  
 <213> Homo sapiens

<400> 2

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Leu | Arg | Val | Phe | Ile | Leu | Tyr | Ala | Glu | Asn | Val | His | Thr | Pro | Asp | 1   | 5   | 10  | 15  |
| Thr | Asp | Ile | Ser | Asp | Ala | Tyr | Cys | Ser | Ala | Val | Phe | Ala | Gly | Val | Lys | 20  | 25  | 30  | 35  |
| Lys | Arg | Thr | Lys | Val | Ile | Lys | Asn | Ser | Val | Asn | Pro | Val | Trp | Asn | Glu | 40  | 45  | 50  | 55  |
| Gly | Phe | Glu | Trp | Asp | Leu | Lys | Gly | Ile | Pro | Leu | Asp | Gln | Gly | Ser | Glu | 60  | 65  | 70  | 75  |
| Leu | His | Val | Val | Val | Lys | Asp | His | Glu | Thr | Met | Gly | Arg | Asn | Arg | Phe | 80  | 85  | 90  | 95  |
| Leu | Gly | Glu | Ala | Lys | Val | Pro | Leu | Arg | Glu | Val | Leu | Ala | Thr | Pro | Ser | 100 | 105 | 110 | 115 |
| Leu | Ser | Ala | Ser | Phe | Asn | Ala | Pro | Leu | Leu | Asp | Thr | Lys | Lys | Gln | Pro | 120 | 125 | 130 | 135 |
| Thr | Gly | Ala | Ser | Leu | Val | Leu | Gln | Val | Ser | Tyr | Thr | Pro | Leu | Pro | Gly | 140 | 145 | 150 | 155 |
| Ala | Val | Pro | Leu | Phe | Pro | Pro | Thr | Pro | Leu | Glu | Gly | Glu | Glu | Asp | Thr | 160 | 165 | 170 | 175 |
| Leu | Pro | Asp | Leu | Asp | Val | Val | Ala | Asp | Thr | Gly | Gly | Glu | Glu | Asp | Thr | 180 | 185 | 190 | 195 |
| Glu | Asp | Gln | Gly | Leu | Thr | Gly | Asp | Glu | Ala | Glu | Pro | Phe | Leu | Asp | Gln | 200 | 205 |     |     |
| Ser | Gly | Gly | Pro | Gly | Ala | Pro | Thr | Thr | Pro | Arg | Lys | Leu | Pro | Ser | Arg |     |     |     |     |
| Pro | Pro | Pro | His | Tyr | Pro | Gly | Ile | Lys | Arg | Lys | Arg | Ser | Ala | Pro | Thr |     |     |     |     |

10/68

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ser | Arg | Lys | Leu | Leu | Ser | Asp | Lys | Pro | Gln | Asp | Phe | Gln | Ile | Arg | Val |
| 210 |     |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| Gln | Val | Ile | Glu | Gly | Arg | Gln | Leu | Pro | Gly | Val | Asn | Ile | Lys | Pro | Val |
| 225 |     |     |     |     | 230 |     |     |     |     | 235 |     |     |     |     | 240 |
| Val | Lys | Val | Thr | Ala | Gly | Gln | Thr | Lys | Arg | Thr | Arg | Ile | His | Lys |     |
|     |     |     |     | 245 |     |     |     | 250 |     |     |     |     | 255 |     |     |
| Gly | Asn | Ser | Pro | Leu | Phe | Asn | Glu | Thr | Leu | Phe | Phe | Asn | Leu | Phe | Asp |
|     |     |     | 260 |     |     |     | 265 |     |     |     |     |     | 270 |     |     |
| Ser | Pro | Gly | Glu | Leu | Phe | Asp | Glu | Pro | Ile | Phe | Ile | Thr | Val | Val | Asp |
|     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
| Ser | Arg | Ser | Leu | Arg | Thr | Asp | Ala | Leu | Leu | Gly | Glu | Phe | Arg | Met | Asp |
|     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
| Val | Gly | Thr | Ile | Tyr | Arg | Glu | Pro | Arg | His | Ala | Tyr | Leu | Arg | Lys | Trp |
| 305 |     |     |     |     | 310 |     |     |     |     | 315 |     |     |     |     | 320 |
| Leu | Leu | Leu | Ser | Asp | Pro | Asp | Asp | Phe | Ser | Ala | Gly | Ala | Arg | Gly | Tyr |
|     |     |     |     | 325 |     |     |     |     | 330 |     |     |     |     | 335 |     |
| Leu | Lys | Thr | Ser | Leu | Cys | Val | Leu | Gly | Pro | Gly | Asp | Glu | Ala | Pro | Leu |
|     |     |     | 340 |     |     |     |     | 345 |     |     |     |     | 350 |     |     |
| Glu | Arg | Lys | Asp | Pro | Ser | Glu | Asp | Lys | Glu | Asp | Ile | Glu | Ser | Asn | Leu |
|     |     | 355 |     |     |     |     | 360 |     |     |     |     | 365 |     |     |     |
| Leu | Arg | Pro | Thr | Gly | Val | Ala | Leu | Arg | Gly | Ala | His | Phe | Cys | Leu | Lys |
|     |     | 370 |     |     |     | 375 |     |     |     |     | 380 |     |     |     |     |
| Val | Phe | Arg | Ala | Glu | Asp | Leu | Pro | Gln | Met | Asp | Asp | Ala | Val | Met | Asp |
| 385 |     |     |     |     | 390 |     |     |     |     | 395 |     |     |     |     | 400 |
| Asn | Val | Lys | Gln | Ile | Phe | Gly | Phe | Glu | Ser | Asn | Lys | Lys | Asn | Leu | Val |
|     |     |     |     | 405 |     |     |     |     | 410 |     |     |     |     | 415 |     |
| Asp | Pro | Phe | Val | Glu | Val | Ser | Phe | Ala | Gly | Lys | Met | Leu | Cys | Ser | Lys |
|     |     |     | 420 |     |     |     |     | 425 |     |     |     |     | 430 |     |     |
| Ile | Leu | Glu | Lys | Thr | Ala | Asn | Pro | Gln | Trp | Asn | Gln | Asn | Ile | Thr | Leu |
|     |     | 435 |     |     |     |     | 440 |     |     |     |     | 445 |     |     |     |
| Pro | Ala | Met | Phe | Pro | Ser | Met | Cys | Glu | Lys | Met | Arg | Ile | Arg | Ile | Ile |
|     | 450 |     |     |     |     | 455 |     |     |     |     | 460 |     |     |     |     |
| Asp | Trp | Asp | Arg | Leu | Thr | His | Asn | Asp | Ile | Val | Ala | Thr | Thr | Tyr | Leu |
| 465 |     |     |     |     | 470 |     |     |     |     | 475 |     |     |     |     | 480 |
| Ser | Met | Ser | Lys | Ile | Ser | Ala | Pro | Gly | Gly | Glu | Ile | Glu | Glu | Glu | Pro |
|     |     |     |     | 485 |     |     |     |     | 490 |     |     |     |     | 495 |     |
| Ala | Gly | Ala | Val | Lys | Pro | Ser | Lys | Ala | Ser | Asp | Leu | Asp | Asp | Tyr | Leu |
|     |     |     | 500 |     |     |     |     | 505 |     |     |     |     | 510 |     |     |
| Gly | Phe | Leu | Pro | Thr | Phe | Gly | Pro | Cys | Tyr | Ile | Asn | Leu | Tyr | Gly | Ser |
|     |     | 515 |     |     |     |     | 520 |     |     |     |     | 525 |     |     |     |
| Pro | Arg | Glu | Phe | Thr | Gly | Phe | Pro | Asp | Pro | Tyr | Thr | Glu | Leu | Asn | Thr |
|     |     | 530 |     |     |     | 535 |     |     |     |     | 540 |     |     |     |     |
| Gly | Lys | Gly | Glu | Gly | Val | Ala | Tyr | Arg | Gly | Arg | Leu | Leu | Leu | Ser | Leu |
| 545 |     |     |     |     | 550 |     |     |     |     | 555 |     |     |     |     | 560 |
| Glu | Thr | Lys | Leu | Val | Glu | His | Ser | Glu | Gln | Lys | Val | Glu | Asp | Leu | Pro |
|     |     |     |     | 565 |     |     |     |     | 570 |     |     |     |     | 575 |     |
| Ala | Asp | Asp | Ile | Leu | Arg | Val | Glu | Lys | Tyr | Leu | Arg | Arg | Arg | Lys | Tyr |
|     |     |     | 580 |     |     |     |     | 585 |     |     |     |     | 590 |     |     |
| Ser | Leu | Phe | Ala | Ala | Phe | Tyr | Ser | Ala | Thr | Met | Leu | Gln | Asp | Val | Asp |
|     |     | 595 |     |     |     |     | 600 |     |     |     |     | 605 |     |     |     |
| Asp | Ala | Ile | Gln | Phe | Glu | Val | Ser | Ile | Gly | Asn | Tyr | Gly | Asn | Lys | Phe |
|     |     | 610 |     |     |     | 615 |     |     |     |     | 620 |     |     |     |     |
| Asp | Met | Thr | Cys | Leu | Pro | Leu | Ala | Ser | Thr | Thr | Gln | Tyr | Ser | Arg | Ala |
| 625 |     |     |     |     | 630 |     |     |     |     | 635 |     |     |     |     | 640 |
| Val | Phe | Asp | Gly | Cys | His | Tyr | Tyr | Tyr | Leu | Pro | Trp | Gly | Asn | Val | Lys |
|     |     |     |     | 645 |     |     |     |     | 650 |     |     |     |     | 655 |     |
| Pro | Val | Val | Val | Leu | Ser | Ser | Tyr | Trp | Glu | Asp | Ile | Ser | His | Arg | Ile |
|     |     |     |     | 660 |     |     |     | 665 |     |     |     |     | 670 |     |     |
| Glu | Thr | Gln | Asn | Gln | Leu | Leu | Gly | Ile | Ala | Asp | Arg | Leu | Glu | Ala | Gly |
|     |     | 675 |     |     |     |     | 680 |     |     |     |     | 685 |     |     |     |
| Leu | Glu | Gln | Val | His | Leu | Ala | Leu | Lys | Ala | Gln | Cys | Ser | Thr | Glu | Asp |
|     | 690 |     |     |     |     | 695 |     |     |     |     | 700 |     |     |     |     |
| Val | Asp | Ser | Leu | Val | Ala | Gln | Leu | Thr | Asp | Glu | Leu | Ile | Ala | Gly | Cys |
| 705 |     |     |     |     | 710 |     |     |     |     | 715 |     |     |     |     | 720 |
| Ser | Gln | Pro | Leu | Gly | Asp | Ile | His | Glu | Thr | Pro | Ser | Ala | Thr | His | Leu |
|     |     |     |     | 725 |     |     |     |     | 730 |     |     |     |     | 735 |     |

11/68

Asp Gln Tyr Leu Tyr Gln Leu Arg Thr His His Leu Ser Gln Ile Thr  
 740 745 750  
 Glu Ala Ala Leu Ala Leu Lys Leu Gly His Ser Glu Leu Pro Ala Ala  
 755 760 765  
 Leu Glu Gln Ala Glu Asp Trp Leu Leu Arg Leu Arg Ala Leu Ala Glu  
 770 775 780  
 Glu Pro Gln Asn Ser Leu Pro Asp Ile Val Ile Trp Met Leu Gln Gly  
 785 790 795 800  
 Asp Lys Arg Val Ala Tyr Gln Arg Val Pro Ala His Gln Val Leu Phe  
 805 810 815  
 Ser Arg Arg Gly Ala Asn Tyr Cys Gly Lys Asn Cys Gly Lys Leu Gln  
 820 825 830  
 Thr Ile Phe Leu Lys Tyr Pro Met Glu Lys Val Pro Gly Ala Arg Met  
 835 840 845  
 Pro Val Gln Ile Arg Val Lys Leu Trp Phe Gly Leu Ser Val Asp Glu  
 850 855 860  
 Lys Glu Phe Asn Gln Phe Ala Glu Gly Lys Leu Ser Val Phe Ala Glu  
 865 870 875 880  
 Thr Tyr Glu Asn Glu Thr Lys Leu Ala Leu Val Gly Asn Trp Gly Thr  
 885 890 895  
 Thr Gly Leu Thr Tyr Pro Lys Phe Ser Asp Val Thr Gly Lys Ile Lys  
 900 905 910  
 Leu Pro Lys Asp Ser Phe Arg Pro Ser Ala Gly Trp Thr Trp Ala Gly  
 915 920 925  
 Asp Trp Phe Val Cys Pro Glu Lys Thr Leu Leu His Asp Met Asp Ala  
 930 935 940  
 Gly His Leu Ser Phe Val Glu Glu Val Phe Glu Asn Gln Thr Arg Leu  
 945 950 955 960  
 Pro Gly Gly Gln Trp Ile Tyr Met Ser Asp Asn Tyr Thr Asp Val Asn  
 965 970 975  
 Gly Glu Lys Val Leu Pro Lys Asp Asp Ile Glu Cys Pro Leu Gly Trp  
 980 985 990  
 Lys Trp Glu Asp Glu Glu Trp Ser Thr Asp Leu Asn Arg Ala Val Asp  
 995 1000 1005  
 Glu Gln Gly Trp Glu Tyr Ser Ile Thr Ile Pro Pro Glu Arg Lys Pro  
 1010 1015 1020  
 Lys His Trp Val Pro Ala Glu Lys Met Tyr Tyr Thr His Arg Arg Arg  
 1025 1030 1035 1040  
 Arg Trp Val Arg Leu Arg Arg Arg Asp Leu Ser Gln Met Glu Ala Leu  
 1045 1050 1055  
 Lys Arg His Arg Gln Ala Glu Ala Glu Gly Glu Gly Trp Glu Tyr Ala  
 1060 1065 1070  
 Ser Leu Phe Gly Trp Lys Phe His Leu Glu Tyr Arg Lys Thr Asp Ala  
 1075 1080 1085  
 Phe Arg Arg Arg Arg Trp Arg Arg Arg Met Glu Pro Leu Glu Lys Thr  
 1090 1095 1100  
 Gly Pro Ala Ala Val Phe Ala Leu Glu Gly Ala Leu Gly Gly Val Met  
 1105 1110 1115 1120  
 Asp Asp Lys Ser Glu Asp Ser Met Ser Val Ser Thr Leu Ser Phe Gly  
 1125 1130 1135  
 Val Asn Arg Pro Thr Ile Ser Cys Ile Phe Asp Tyr Gly Asn Arg Tyr  
 1140 1145 1150  
 His Leu Arg Cys Tyr Met Tyr Gln Ala Arg Asp Leu Ala Ala Met Asp  
 1155 1160 1165  
 Lys Asp Ser Phe Ser Asp Pro Tyr Ala Ile Val Ser Phe Leu His Gln  
 1170 1175 1180  
 Ser Gln Lys Thr Val Val Val Lys Asn Thr Leu Asn Pro Thr Trp Asp  
 1185 1190 1195 1200  
 Gln Thr Leu Ile Phe Tyr Glu Ile Glu Ile Phe Gly Glu Pro Ala Thr  
 1205 1210 1215  
 Val Ala Glu Gln Pro Pro Ser Ile Val Val Glu Leu Tyr Asp His Asp  
 1220 1225 1230  
 Thr Tyr Gly Ala Asp Glu Phe Met Gly Arg Cys Ile Cys Gln Pro Ser  
 1235 1240 1245  
 Leu Glu Arg Met Pro Arg Leu Ala Trp Phe Pro Leu Thr Arg Gly Ser  
 1250 1255 1260

12/68

Gln Pro Ser Gly Glu Leu Leu Ala Ser Phe Glu Leu Ile Gln Arg Glu  
 1265 1270 1275 1280  
 Lys Pro Ala Ile His His Ile Pro Gly Phe Glu Val Gln Glu Thr Ser  
 1285 1290 1295  
 Arg Ile Leu Asp Glu Ser Glu Asp Thr Asp Leu Pro Tyr Pro Pro Pro  
 1300 1305 1310  
 Gln Arg Glu Ala Asn Ile Tyr Met Val Pro Gln Asn Ile Lys Pro Ala  
 1315 1320 1325  
 Leu Gln Arg Thr Ala Ile Glu Ile Leu Ala Trp Gly Leu Arg Asn Met  
 1330 1335 1340  
 Lys Ser Tyr Gln Leu Ala Asn Ile Ser Ser Pro Ser Leu Val Val Glu  
 1345 1350 1355 1360  
 Cys Gly Gly Gln Thr Val Gln Ser Cys Val Ile Arg Asn Leu Arg Lys  
 1365 1370 1375  
 Asn Pro Asn Phe Asp Ile Cys Thr Leu Phe Met Glu Val Met Leu Pro  
 1380 1385 1390  
 Arg Glu Glu Leu Tyr Cys Pro Pro Ile Thr Val Lys Val Ile Asp Asn  
 1395 1400 1405  
 Arg Gln Phe Gly Arg Arg Pro Val Val Gly Gln Cys Thr Ile Arg Ser  
 1410 1415 1420  
 Leu Glu Ser Phe Leu Cys Asp Pro Tyr Ser Ala Glu Ser Pro Ser Pro  
 1425 1430 1435 1440  
 Gln Gly Gly Pro Asp Asp Val Ser Leu Leu Ser Pro Gly Glu Asp Val  
 1445 1450 1455  
 Leu Ile Asp Ile Asp Asp Lys Glu Pro Leu Ile Pro Ile Gln Glu Glu  
 1460 1465 1470  
 Glu Phe Ile Asp Trp Trp Ser Lys Phe Phe Ala Ser Ile Gly Glu Arg  
 1475 1480 1485  
 Glu Lys Cys Gly Ser Tyr Leu Glu Lys Asp Phe Asp Thr Leu Lys Val  
 1490 1495 1500  
 Tyr Asp Thr Gln Leu Glu Asn Val Glu Ala Phe Glu Gly Leu Ser Asp  
 1505 1510 1515 1520  
 Phe Cys Asn Thr Phe Lys Leu Tyr Arg Gly Lys Thr Gln Glu Glu Thr  
 1525 1530 1535  
 Glu Asp Pro Ser Val Ile Gly Glu Phe Lys Gly Leu Phe Lys Ile Tyr  
 1540 1545 1550  
 Pro Leu Pro Glu Asp Pro Ala Ile Pro Met Pro Pro Arg Gln Phe His  
 1555 1560 1565  
 Gln Leu Ala Ala Gln Gly Pro Gln Glu Cys Leu Val Arg Ile Tyr Ile  
 1570 1575 1580  
 Val Arg Ala Phe Gly Leu Gln Pro Lys Asp Pro Asn Gly Lys Cys Asp  
 1585 1590 1595 1600  
 Pro Tyr Ile Lys Ile Ser Ile Gly Lys Lys Ser Val Ser Asp Gln Asp  
 1605 1610 1615  
 Asn Tyr Ile Pro Cys Thr Leu Glu Pro Val Phe Gly Lys Met Phe Glu  
 1620 1625 1630  
 Leu Thr Cys Thr Leu Pro Leu Glu Lys Asp Leu Lys Ile Thr Leu Tyr  
 1635 1640 1645  
 Asp Tyr Asp Leu Leu Ser Lys Asp Glu Lys Ile Gly Glu Thr Val Val  
 1650 1655 1660  
 Asp Leu Glu Asn Arg Leu Leu Ser Lys Phe Gly Ala Arg Cys Gly Leu  
 1665 1670 1675 1680  
 Pro Gln Thr Tyr Cys Val Ser Gly Pro Asn Gln Trp Arg Asp Gln Leu  
 1685 1690 1695  
 Arg Pro Ser Gln Leu Leu His Leu Phe Cys Gln Gln His Arg Val Lys  
 1700 1705 1710  
 Ala Pro Val Tyr Arg Thr Asp Arg Val Met Phe Gln Asp Lys Glu Tyr  
 1715 1720 1725  
 Ser Ile Glu Glu Ile Glu Ala Gly Arg Ile Pro Asn Pro His Leu Gly  
 1730 1735 1740  
 Pro Val Glu Glu Arg Leu Ala Leu His Val Leu Gln Gln Gln Gly Leu  
 1745 1750 1755 1760  
 Val Pro Glu His Val Glu Ser Arg Pro Leu Tyr Ser Pro Leu Gln Pro  
 1765 1770 1775  
 Asp Ile Glu Gln Gly Lys Leu Gln Met Trp Val Asp Leu Phe Pro Lys  
 1780 1785 1790

13/68

Ala Leu Gly Arg Pro Gly Pro Pro Phe Asn Ile Thr Pro Arg Arg Ala  
 1795 1800 1805  
 Arg Arg Phe Phe Leu Arg Cys Ile Ile Trp Asn Thr Arg Asp Val Ile  
 1810 1815 1820  
 Leu Asp Asp Leu Ser Leu Thr Gly Glu Lys Met Ser Asp Ile Tyr Val  
 1825 1830 1835 1840  
 Lys Gly Trp Met Ile Gly Phe Glu Glu His Lys Gln Lys Thr Asp Val  
 1845 1850 1855  
 His Tyr Arg Ser Leu Gly Gly Glu Gly Asn Phe Asn Trp Arg Phe Ile  
 1860 1865 1870  
 Phe Pro Phe Asp Tyr Leu Pro Ala Glu Gln Val Cys Thr Ile Ala Lys  
 1875 1880 1885  
 Lys Asp Ala Phe Trp Arg Leu Asp Lys Thr Glu Ser Lys Ile Pro Ala  
 1890 1895 1900  
 Arg Val Val Phe Gln Ile Trp Asp Asn Asp Lys Phe Ser Phe Asp Asp  
 1905 1910 1915 1920  
 Phe Leu Gly Ser Leu Gln Leu Asp Leu Asn Arg Met Pro Lys Pro Ala  
 1925 1930 1935  
 Lys Thr Ala Lys Lys Cys Ser Leu Asp Gln Leu Asp Asp Ala Phe His  
 1940 1945 1950  
 Pro Glu Trp Phe Val Ser Leu Phe Glu Gln Lys Thr Val Lys Gly Trp  
 1955 1960 1965  
 Trp Pro Cys Val Ala Glu Glu Gly Glu Lys Lys Ile Leu Ala Gly Lys  
 1970 1975 1980  
 Leu Glu Met Thr Leu Glu Ile Val Ala Glu Ser Glu His Glu Glu Arg  
 1985 1990 1995 2000  
 Pro Ala Gly Gln Gly Arg Asp Glu Pro Asn Met Asn Pro Lys Leu Glu  
 2005 2010 2015  
 Asp Pro Arg Arg Pro Asp Thr Ser Phe Leu Trp Phe Thr Ser Pro Tyr  
 2020 2025 2030  
 Lys Thr Met Lys Phe Ile Leu Trp Arg Arg Phe Arg Trp Ala Ile Ile  
 2035 2040 2045  
 Leu Phe Ile Ile Leu Phe Ile Leu Leu Leu Phe Leu Ala Ile Phe Ile  
 2050 2055 2060  
 Tyr Ala Phe Pro Asn Tyr Ala Ala Met Lys Leu Val Lys Pro Phe Ser  
 2065 2070 2075 2080

<210> 3  
 <211> 5915  
 <212> DNA  
 <213> Homo sapiens

<400> 3  
 tcgaccgccc agccagggtgc aaaatgccgt gtcattggga gactccgcag ccggagcatt 60  
 agattacagc tcgacgggagc tcgggaaggg cggcgggggg ggaagatgag cagaagcccc 120  
 tgttctcgga acgcccgtcg acaagcgggg tgagcgcagg cggggcgggg acccagccta 180  
 gccactgga gcagccgggg gtggcccgtt cccctttaag agcaactgct ctaagccagg 240  
 agccagagat tcgagccggc ctgcgccagc cagccctctc cagcgagggg acccacaagc 300  
 ggcgcctcgg ccctcccgac ctttccgagc cctctttgcg ccctgggcgc acgggggccct 360  
 acacgcgcca agcatgctga ggtcttcat cctctatgcc gagaacgtcc acacacccga 420  
 caccgacatc agcgtgcct actgctccgc ggtgtttgca ggggtgaaga agagaaccaa 480  
 agtcatcaag aacagcgtga accctgtatg gaatgaggga tttgaatggg acctcaaggg 540  
 catccccctg gaccagggtc ctgagcttca tgtggtggtc aaagaccatg agacgatggg 600  
 gaggaacagg ttcttggggg aagccaaggt cccactccga gaggtcctcg ccaccctag 660  
 tctgtccgcc agcttcaatg ccccttgct ggacaccaag aagcagccca caggggcctc 720  
 gctgggtcctg caggtgtcct acacaccgct gcctggagct gtgcccctgt tcccgcccc 780  
 tactcctctg gagccctccc cgactctgcc tgacctggat gtagtggcag acacaggagg 840  
 agaggaagac acagaggacc agggactcac tggagatgag gcggagccat tcttgatca 900  
 aagcggaggc ccggggggctc ccaccacccc aaggaaacta ccttcacgtc ctccgcccc 960  
 ctaccccggg atcaaaagaa agcgaagtgc gcctacatct agaaagctgc tgtcagacaa 1020  
 accgcaggat ttccagatca ggggtccagg gatcgagggg cgccagctgc cgggggtgaa 1080  
 catcaagcct gtggtcaagg ttaccgctgc agggcagacc aagcggagcg ggatccacaa 1140  
 gggaaacagc ccactcttca atgagactct tttcttcaac ttgtttgact ctctgggga 1200  
 gctgtttgat gagcccatct ttatcacggt ggtagactct cgttctctca ggacagatgc 1260  
 tctcctcggg gagttccgga tggacgtggg caccatttac agagagcccc ggcacgccta 1320  
 tctcaggaag tggctgctgc tctcagaccc tgatgacttc tctgctgggg ccagaggcta 1380  
 cctgaaaaca agcctttgtg tgctgggggccc tggggacgaa gcgcctctgg agagaaaaga 1440

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| ccccctctgaa | gacaaggagg  | acattgaaag  | caacctgctc  | cgccccacag  | gcgtagccct  | 1500 |
| gcgaggagcc  | cacttctgcc  | tgaaggtctt  | ccgggcccag  | gacttgccgc  | agatggacga  | 1560 |
| tgccgtgatg  | gacaacgtga  | aacagatctt  | tggcttcgag  | agtaacaaga  | agaacttggt  | 1620 |
| ggaccccttt  | gtggaggtca  | gctttgcggg  | gaaaatgctg  | tgcaacaaga  | tcttgagaga  | 1680 |
| gacggccaac  | cctcagtggg  | accagaacat  | cacactgcct  | gccatgtttc  | cctccatgtg  | 1740 |
| cgaaaaaatg  | aggattcgta  | tcatagactg  | ggaccgcctg  | actcacaatg  | acatcgtggc  | 1800 |
| taccacctac  | ctgagtatgt  | cgaaaatctc  | tgccccctgga | ggagaaatag  | aagaggagcc  | 1860 |
| tgcaggtgct  | gtcaagcctt  | cgaaagcctc  | agacttggtat | gactacctgg  | gcttcctccc  | 1920 |
| cacttttggg  | ccctgctaca  | tcaacctcta  | tggcagtcct  | agagagttca  | caggcttccc  | 1980 |
| agacccctac  | acagagctca  | acacaggcaa  | gggggaaggt  | gtggccttat  | gtggcggct   | 2040 |
| tctgctctcc  | ctggagacca  | agctggtgga  | gcacagtga   | cagaaggtgg  | aggaccttcc  | 2100 |
| tgcggtatgac | atcctccggg  | tggagaagta  | ccttaggagg  | cgcaagtact  | ccctgtttgc  | 2160 |
| ggccttctac  | tcagccacca  | tgctgcagga  | tgtggatgat  | gccatccagt  | ttgaggtcag  | 2220 |
| catcggaac   | tacgggaaca  | agttcgacat  | gacctgcctg  | ccgctggcct  | ccaccactca  | 2280 |
| gtacagccgt  | gcagtccttg  | acgggtgcc   | ctactactac  | ctaccctggg  | gtaacgtgaa  | 2340 |
| acctgtggtg  | ctgctgtcat  | cctactggga  | ggacatcagc  | catagaatcg  | agactcagaa  | 2400 |
| ccagctgctt  | gggattgctg  | accggctgga  | agctggcctg  | gagcaggtcc  | acctggccct  | 2460 |
| gaaggcgag   | tgctccacgg  | aggacgtgga  | ctcgtggtg   | gctcagctga  | cggatgagct  | 2520 |
| catcgaggg   | tgcagccagc  | ctctgggtga  | catccatgag  | acaccctctg  | ccaccacct   | 2580 |
| ggaccagtac  | ctgtaccagc  | tgcgcacca   | tcacctgagc  | caaactcactg | aggctgcct   | 2640 |
| ggccctgaag  | ctcggccaca  | gtgagctccc  | tgcagctctg  | gagcaggcgg  | aggactggct  | 2700 |
| cctgctctg   | cgtgccctgg  | cagaggagcc  | ccagaacagc  | ctgccggaca  | tcgtcatctg  | 2760 |
| gatgctgcag  | ggagacaagc  | gtgtggcata  | ccagcgggtg  | cccgccacc   | aagtccctct  | 2820 |
| ctccggcg    | ggtgccaaact | actgtggcaa  | gaattgtggg  | aagctacaga  | caatctttct  | 2880 |
| gaaatatccg  | atggagaagg  | tgcctggcgc  | ccggatgcca  | gtgcagatac  | gggtcaggct  | 2940 |
| gtggtttggg  | ctctctgtgg  | atgagaagga  | gttcaaccag  | tttgctgagg  | ggaagctgtc  | 3000 |
| tgtctttgct  | gaaacctatg  | agaacgagac  | taagtggcc   | cttggtggga  | actggggcac  | 3060 |
| aacgggcctc  | acctacccca  | agttttctga  | cgtaacgggc  | aagatcaagc  | tacccaagga  | 3120 |
| cagcttccgc  | ccctcgcccg  | gctggacctg  | ggctggagat  | tggttcgtgt  | gtccggagaa  | 3180 |
| gactctgctc  | catgacatgg  | acgcccgttc  | cctgagcttc  | gtggaagagg  | tggttgagaa  | 3240 |
| ccagaccgg   | cttcccggag  | gccagtggtg  | ctacatgagt  | gacaactaca  | ccgatgtgaa  | 3300 |
| cggggagaa   | gtgcttccca  | aggatgacat  | tgagtggcca  | ctgggctgga  | agtgggaaga  | 3360 |
| tgaggaatgg  | tccacagacc  | tcaaccgggc  | tgtcgatgag  | caaggctggg  | agtatagcat  | 3420 |
| caccatcccc  | ccggagcgga  | agccgaagca  | ctgggtccct  | gctgagaaga  | tgtactacac  | 3480 |
| acaccgacgg  | ggcgctggg   | tgcgcctgcg  | caggagggtg  | ctcagccaaa  | tggaaagcact | 3540 |
| gaaaaggcac  | aggcaggcgg  | aggcggaggg  | cgagggtctg  | gagtacgcct  | ctcttttttg  | 3600 |
| ctggaagtcc  | cacctcgagt  | accgcaagac  | agatgccttc  | cgccgccgcc  | gctggcgccg  | 3660 |
| tcgcatggag  | ccactggaga  | agacggggcc  | tgcagctgtg  | tttgcccttg  | agggggccct  | 3720 |
| gggcggcgtg  | atggatgaca  | agagtgaaga  | ttccatgtcc  | gtctccacct  | tgagcttcgg  | 3780 |
| tgtgaacaga  | cccacgattt  | cctgcatatt  | cgactatggg  | aaccgctacc  | atctacgctg  | 3840 |
| ctacatgtac  | caggcccggg  | acctggctgc  | gatggacaag  | gactcttttt  | ctgatcccta  | 3900 |
| tgccatcgtc  | tccttcctgc  | accagagcca  | gaagacgggt  | gtggtgaaga  | acacccttaa  | 3960 |
| ccccacctgg  | gaccagacgc  | tcatcttcta  | cgagatcgag  | atctttggcg  | agccggccac  | 4020 |
| agttgctgag  | caaccgcccc  | gcatttgtgt  | ggagctgtac  | gaccatgaca  | cttatggtgc  | 4080 |
| agacgagttt  | atgggtcgct  | gcattgtgca  | accgagctct  | gaacggatgc  | cacgggtggc  | 4140 |
| ctggttccca  | ctgacgaggg  | gcagccagcc  | gtcgggggag  | ctgctggcct  | cttttgagct  | 4200 |
| catccagaga  | gagaagccgg  | ccatccacca  | tattcctggt  | tttgaggtgc  | aggagacatc  | 4260 |
| aaggatcctg  | gatgagtctg  | aggacacaga  | cctgccctac  | ccaccacccc  | agagggaggc  | 4320 |
| caacatctac  | atggttcctc  | agaacatcaa  | gccagcgctc  | cagcgtaccg  | ccatcgagat  | 4380 |
| cctggcatgg  | ggcctgcgga  | acatgaagag  | ttaccagctg  | gccaacatct  | cctccccag   | 4440 |
| cctcgtggta  | gagtgtgggg  | gccagacggg  | gcagtcctgt  | gtcatcagga  | acctccggaa  | 4500 |
| gaaccccaac  | tttgacatct  | gcacctctt   | catggaagtg  | atgctgcccc  | gggaggagct  | 4560 |
| ctactgcccc  | cccatcacgg  | tcaaggtcat  | cgataaccgc  | cagtttgggc  | gccggcctgt  | 4620 |
| ggtgggcccag | tgtaccatcc  | gctccctgga  | gagcttcctg  | tgtgacctct  | actcggcgga  | 4680 |
| gagtcacatc  | ccacagggtg  | gcccagacga  | tgtgagccta  | ctcagtcctg  | gggaagacgt  | 4740 |
| gctcatcgac  | attgatgaca  | aggagcccct  | catccccatc  | caggaggaag  | agttcatcga  | 4800 |
| ttggtggagc  | aaattctttg  | cctccatagg  | ggagaggga   | aagtgcggct  | cctacctgga  | 4860 |
| gaaggatttt  | gacaccctga  | aggtctatga  | cacacagctg  | gagaatgtgg  | aggcctttga  | 4920 |
| gggctgtctc  | gactttttgt  | acaccttcaa  | gctgtaccgg  | ggcaagacgc  | aggagagac   | 4980 |
| agaagatcca  | tctgtgattg  | gtgaatttaa  | gggcctcttc  | aaaattttatc | ccctcccaga  | 5040 |
| agacccagcc  | atccccatgc  | ccccaaagaca | gttccaccag  | ctggccgccc  | agggacccca  | 5100 |
| ggagtgcctg  | gtccgtatct  | acattgtccg  | agcatttgcc  | ctgcagccca  | aggaccccaa  | 5160 |
| tggaaagtgt  | gacccctaca  | tcaagatctc  | catagggaag  | aaatcagtga  | gtgaccagga  | 5220 |
| taactacatc  | cctgcacgc   | tggagccctg  | atttggaaag  | atgttcgagc  | tgacctgcac  | 5280 |
| tctgcctctg  | gagaaggacc  | taaagatcac  | tctctatgac  | tatgacctcc  | tctccaagga  | 5340 |
| cgaaaagatc  | ggtgagacgg  | tcgtcgacct  | ggagaacagg  | ctgctgtcca  | agtttggggc  | 5400 |
| tcgctgtgga  | ctcccacaga  | cctactgtgt  | ctctggaccg  | aaccagtggc  | gggaccagct  | 5460 |



15/68

|             |            |            |            |            |            |      |
|-------------|------------|------------|------------|------------|------------|------|
| ccgccccctcc | cagctcctcc | acctcttctg | ccagcagcat | agagtcaagg | cacctgtgta | 5520 |
| ccggacagac  | cgtgtaatgt | ttcaggataa | agaatattcc | attgaagaga | tagaggctgg | 5580 |
| caggatccca  | aaccacacac | tgggcccagt | ggaggagcgt | ctggctctgc | atgtgcttca | 5640 |
| gcagcagggc  | ctggtccccg | agcacgtgga | gtcacggccc | ctctacagcc | ccctgcagcc | 5700 |
| agacatcgag  | caggggaagc | tgcagatgtg | ggtcgacctc | tttccgaagg | ccctggggcg | 5760 |
| gcctggacct  | cccttcaaca | tcacccacg  | gagagccaga | agggttttcc | tgcgttgat  | 5820 |
| tatctggaat  | accagagatg | tgatcctgga | tgacctgagc | ctcacggggg | agaagatgag | 5880 |
| cgacatttat  | gtgaaagggt | ggatgattgg | ctttg      |            |            | 5915 |

<210> 4  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 4  
 tgggacctca aagggcatcc 20

<210> 5  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 5  
 accatgctgt aggatgtgga 20

<210> 6  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 6  
 gggaggtgaa gcaacttcaa 20

<210> 7  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 7  
 ctcacgggggt agaagatgag 20

<210> 8  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 8  
 cagggccgag atgagcccaa 20

<210> 9  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 9  
 acatcaaggg tcctggatga 20

<210> 10  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 10  
 ctgtggcggt gtttccggtg 20

<210> 11  
 <211> 20

16/68

<212> DNA  
<213> Homo sapiens

<400> 11  
acagacgtgc gttatcgcttc 20

<210> 12  
<211> 20  
<212> DNA  
<213> Homo sapiens

<400> 12  
aagactgagc aaaatcccag 20

<210> 13  
<211> 6912  
<212> DNA  
<213> Homo sapiens

<400> 13

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| tcgaccgccc  | agccagggtgc | aaaatgccgt  | gtcattggga  | gactccgcag  | ccggagcatt  | 60   |
| agattacagc  | tcgacggagc  | tcgggaaggg  | cgccgggggt  | ggaagatgag  | cagaagcccc  | 120  |
| tggttcgga   | acgccggctg  | acaagcgggg  | tgagcgcagg  | cggggcgggg  | acccagccta  | 180  |
| gcccactgga  | gcagccgggg  | gtggcccgtt  | cccccttaag  | agcaactgct  | ctaagccagg  | 240  |
| agccagagat  | tcgagccggc  | ctcgcccagc  | cagccctctc  | cagcgagggg  | acccacaagc  | 300  |
| ggcgccctcg  | ccctcccagc  | ctttccgagc  | cctctttgcg  | ccctgggcgc  | acggggccct  | 360  |
| acacgcgcca  | agcatgctga  | gggtcttcat  | cctctatgcc  | gagaacgtcc  | acacaccgga  | 420  |
| caccgacatc  | agcgatgcct  | actgctccgc  | ggtgtttgca  | ggggtgaaga  | agagaaccaa  | 480  |
| agtcatacaag | aacagcgtga  | accctgtatg  | gaatgagggg  | tttgaatggg  | acctcaaagg  | 540  |
| gcatccccct  | ggaccagggc  | tctgagcttc  | atgtgggtgt  | caaagaccat  | gagacgatgg  | 600  |
| ggaggaacag  | gttcctgggg  | gaagccaagg  | tcccactccg  | agaggtcctc  | gccaccctta  | 660  |
| gtctgtccgc  | cagcttcaat  | gccccctgc   | tggacaccaa  | gaagcagccc  | acaggggcct  | 720  |
| cgctggctct  | gcaggtgtcc  | tacacaccgc  | tgccctggagc | tggtccccctg | ttcccgcctc  | 780  |
| ctactcctct  | ggagccctcc  | ccgactctgc  | ctgacctgga  | tgtagtgga   | gacacaggag  | 840  |
| gagaggaaga  | cacagaggac  | cagggactca  | ctggagatga  | ggcgagacca  | ttcctggatc  | 900  |
| aaagcggagg  | cccgggggct  | cccaccacc   | caaggaaact  | accttcacgt  | cctccgcccc  | 960  |
| actaccccgg  | gatcaaaaaga | aagcgaagtg  | cgcctacatc  | tagaaaagctg | ctgtcagaca  | 1020 |
| aaccgcagga  | tttccagatc  | aggggtccagg | tgatcgaggg  | gcgccagctg  | ccgggggtga  | 1080 |
| acatcaagcc  | tgtggtcaag  | gttaccgctg  | cagggcagac  | caagcggacg  | cggatccaca  | 1140 |
| agggaaacag  | cccactcttc  | aatgagactc  | ttttcttcaa  | cttgtttgac  | tctcctgggg  | 1200 |
| agctgtttga  | tgagcccatc  | tttatcaccg  | tggtagactc  | tcgttctctc  | aggacagatg  | 1260 |
| ctctcctcgg  | ggagttccgg  | atggacgtgg  | gcaccattta  | cagagagccc  | cggcacgcct  | 1320 |
| atctcaggaa  | gtggctgctg  | ctctcagacc  | ctgatgactt  | ctctgctggg  | gccagaggct  | 1380 |
| acctgaaaac  | aagcctttgt  | gtgctggggc  | ctggggacga  | agcgcctctg  | gagagaaaag  | 1440 |
| acccctctga  | agacaaggag  | gacattgaaa  | gcaacctgct  | ccggccccaca | ggcgtagccc  | 1500 |
| tgcgaggagc  | ccacttctgc  | ctgaaggctc  | tccgggccga  | ggacttgccg  | cagatggacg  | 1560 |
| atgccgtgat  | ggacaacgtg  | aaacagatct  | ttggcttcga  | gagtaacaag  | aagaacttgg  | 1620 |
| tggacccctt  | tgtggaggtc  | agctttgcgg  | ggaaaatgct  | gtgcagcaag  | atcttgagga  | 1680 |
| agacggccaa  | ccctcagttg  | aaccagaaca  | tcacactgcc  | tgccatgttt  | ccctccatgt  | 1740 |
| gcgaaaaaat  | gaggattcgt  | atcatagact  | gggaccgcct  | gactcacaat  | gacatcgtgg  | 1800 |
| ctaccaccta  | cctgagtatg  | tcgaaaatct  | ctgcccctgg  | aggagaaata  | gaagaggagc  | 1860 |
| ctgcaggtgc  | tgtcaagcct  | tcgaaagcct  | cagacttgga  | tgactacctg  | ggcttccctc  | 1920 |
| ccacttttgg  | gccctgctac  | atcaacctct  | atggcagctc  | cagagagttc  | acaggcttcc  | 1980 |
| cagaccctta  | cacagagctc  | aacacaggga  | agggggaagg  | tgtggcttat  | cgtggccggc  | 2040 |
| ttctgctctc  | cctggagacc  | aagctggtgg  | agcacagtga  | acagaagggtg | gaggaccttc  | 2100 |
| ctgcggatga  | catcctccgg  | gtggagaagt  | accttaggag  | gcgcaagtac  | tcctgttttg  | 2160 |
| cggccttcta  | ctcagccacc  | atgctgtagg  | atgtggatga  | tgccatccag  | tttgagggtca | 2220 |
| gcatcggaag  | ctacgggaac  | aagttcgaca  | tgacctgcct  | gccgctggcc  | tcaccactc   | 2280 |
| agtacagccg  | tgcagctctt  | gacgggtgcc  | actactacta  | cctaccctgg  | gctaacttga  | 2340 |
| aacctgtggt  | ggtgctgtca  | tcctactggg  | aggacatcag  | ccatagaatc  | gagactcaga  | 2400 |
| accagctgct  | tgggattgct  | gaccggctgg  | aagctggcct  | ggagcaggtc  | cacctggccc  | 2460 |
| tgaaggcgca  | gtgctccacg  | gaggacgtgg  | actcgctggg  | ggctcagctg  | acggatgagc  | 2520 |
| tcacgcagag  | ctgcagccag  | cctctgggtg  | acatccatga  | gacaccctct  | gccaccacc   | 2580 |
| tggaccagta  | cctgtaccag  | ctgcgcaccc  | atcacctgag  | ccaaatcact  | gaggctgccc  | 2640 |
| tggccctgaa  | cctcgccac   | agtgaagctc  | ctgcagctct  | ggagcagggc  | gaggactggc  | 2700 |
| tcctgctgct  | gcgtgccctg  | gcagaggagc  | cccagaacag  | cctgccggac  | atcgtcatct  | 2760 |
| ggatgctgca  | gggagacaag  | cgtgtggcat  | accagcgggt  | gcccgccac   | caagtcctct  | 2820 |

|             |             |             |             |            |             |      |
|-------------|-------------|-------------|-------------|------------|-------------|------|
| tctcccgccg  | gggtgccaac  | tactgtggca  | agaattgtgg  | gaagctacag | acaatctttc  | 2880 |
| tgaaatatcc  | gatggagaag  | gtgcctggcg  | cccggatgcc  | agtgcagata | cgggtcaagc  | 2940 |
| tgtggtttgg  | gctctctgtg  | gatgagaagg  | agttcaacca  | gtttgctgag | gggaagctgt  | 3000 |
| ctgtctttgc  | tgaaacctat  | gagaacgaga  | ctaagttggc  | ccttgttggg | aactggggca  | 3060 |
| caacgggcct  | cacctacccc  | aagttttctg  | acgtcacggg  | caagatcaag | ctacccaagg  | 3120 |
| acagcttcgg  | ccccctggcc  | ggctggacct  | gggctggaga  | ttggttcgtg | tgtccggaga  | 3180 |
| agactctgct  | ccatgacatg  | gacgccggtc  | acctgagctt  | cgtggaagag | gtgtttgaga  | 3240 |
| accagaccgg  | gcttcccggg  | ggccagtggg  | tctacatgag  | tgacaactac | accgatgtga  | 3300 |
| acggggagaa  | ggtgcttccc  | aaggatgaca  | ttgagtggcc  | actgggctgg | aagtgggaag  | 3360 |
| atgaggaatg  | gtccacagac  | ctcaaccggg  | ctgtogatga  | gcaaggctgg | gagtatagca  | 3420 |
| tcaccatccc  | cccggagcgg  | aagccgaagc  | actgggtccc  | tgctgagaag | atgtactaca  | 3480 |
| cacaccgacg  | gcggcgctgg  | gtgcgcctgc  | gcaggaggga  | tctcagccaa | atggaagcac  | 3540 |
| tgaaaaggca  | caggcaggcg  | gaggcggagg  | gcgagggctg  | ggagtacgcc | tctctttttg  | 3600 |
| gctggaagtt  | ccacctcgag  | taccgcaaga  | cagatgcctt  | ccgccgcgcg | cgctggcgcc  | 3660 |
| gtcgcattga  | gccactggag  | aagacggggc  | ctgcagctgt  | gtttgccctt | gagggggccc  | 3720 |
| tgggcgcgct  | gatggatgac  | aagagtgaag  | attccatgtc  | cgtctccacc | ttgagcttcg  | 3780 |
| gtgtgaacag  | accacgatt   | tcctgcatat  | tcgactatgg  | gaaccgctac | catctacgct  | 3840 |
| gctacatgta  | ccaggcccgg  | gacctggctg  | cgatggacaa  | ggactctttt | tctgatccct  | 3900 |
| atgccatcgt  | ctccttcctg  | caccagagcc  | agaagacggt  | ggtggtgaag | aacaccttta  | 3960 |
| accccacctg  | ggaccagacg  | ctcatcttct  | acgagatcga  | gatctttggc | gagccggcca  | 4020 |
| cagttgctga  | gcaaccgccc  | agcattgtgg  | tggagctgta  | cgaccatgac | acttatgggtg | 4080 |
| cagacgagtt  | tatgggtcgc  | tgcattctgtc | aaccgagctt  | ggaacggatg | ccacggctgg  | 4140 |
| cctggttccc  | actgacgagg  | ggcagccagc  | cgctcgggga  | gctgctggcc | tcttttgagc  | 4200 |
| tcatccagag  | agagaagccg  | gccatccacc  | atatctctgg  | ttttgaggtg | caggagacat  | 4260 |
| caaggatcct  | ggatgagtct  | gaggacacag  | acctgcccta  | cccaccaccc | cagagggagg  | 4320 |
| ccaacatcta  | catggttccct | cagaacatca  | agccagcgct  | ccagcgtacc | gccatcgaga  | 4380 |
| tectggcatg  | gggcctgcgg  | aacatgaaga  | gttaccagct  | ggccaacatc | tcctccccc   | 4440 |
| gcctcgtggt  | agagtgtggg  | ggccagacgg  | tgcagtcctg  | tgtcatcagg | aacctccgga  | 4500 |
| agaaccccaa  | ctttgacatc  | tgcacctctc  | tcattggaagt | gatgctggcc | agggaggagc  | 4560 |
| tctactgccc  | ccccatcacc  | gtcaaggcca  | tcgataaccg  | ccagtttggc | cgccggcctg  | 4620 |
| tgggtggcca  | gtgtaccatc  | cgctccctgg  | agagcttccct | gtgtgacccc | tactcgccgg  | 4680 |
| agagtccatc  | cccacagggt  | ggcccagacg  | atgtgagcct  | actcagtcct | ggggaagacg  | 4740 |
| tgctcatcga  | cattgatgac  | aaggagcccc  | tcattcccat  | ccaggaggaa | gagttcatcg  | 4800 |
| attggtggag  | caaattcttt  | gcctccatag  | gggagaggga  | aaagtgcggc | tcctacctgg  | 4860 |
| agaaggattt  | tgacacctg   | aaggtctatg  | acacacagct  | ggagaatgtg | gagccctttg  | 4920 |
| agggcctgtc  | tgacttttgt  | aacaccttca  | agctgtaccg  | gggcaagacg | caggaggaga  | 4980 |
| cagaagatcc  | atctgtgatt  | ggtgaattta  | agggcctctt  | caaaatztat | ccccctccag  | 5040 |
| aagaccacag  | catccccatg  | cccccaagac  | agttccacca  | gctggccgcc | cagggacccc  | 5100 |
| aggagtgcct  | ggtccgtatc  | tacattgtcc  | gagcatttgg  | cctgcagccc | aaggacccca  | 5160 |
| atggaaagtg  | tgatctttac  | atcaagatct  | ccatagggaa  | gaaatcagtg | agtgaccagg  | 5220 |
| ataactacat  | ccccctgcacg | ctggagcccc  | tatttggaag  | gatgttcgag | ctgacctgca  | 5280 |
| ctctgcctct  | ggagaaggac  | ctaaagatca  | ctctctatga  | ctatgacctc | ctctccaagg  | 5340 |
| acgaaaagat  | cggtgagacg  | gtcgtcgacc  | tggagaacag  | gctgctgtcc | aagtttgggg  | 5400 |
| ctcgtgtgtg  | actcccacag  | acctactgtg  | tctctggacc  | gaaccagtgg | cgggaccagc  | 5460 |
| tccgcccctc  | ccagctcttc  | cacctcttct  | gccagcagca  | tagagtcaag | gcacctgtgt  | 5520 |
| accggacaga  | ccgtgtaatg  | tttcaggata  | aagaatattc  | cattgaagag | atagaggctg  | 5580 |
| gcaggatccc  | aaacccacac  | ctgggcccag  | tggaggagcg  | tctggctctg | catgtgcttc  | 5640 |
| agcagcaggg  | cctggtcccc  | gagcacgtgg  | agtcacggcc  | cctctacagc | ccccctgcagc | 5700 |
| cagacatcga  | gcagggggaag | ctgcagatgt  | gggtcgacct  | atttccgaag | cgccctggggc | 5760 |
| ggcctggacc  | tcccttcaac  | atcaccccac  | ggagagccag  | aaggtttttc | ctgcgttgta  | 5820 |
| ttatctggaa  | taccagagat  | gtgatcctgg  | atgacctgag  | cctcacgggg | gagaagatga  | 5880 |
| gcgacattta  | tgtgaaagggt | tggatgattg  | gctttgaaga  | acacaagcaa | aagacagacg  | 5940 |
| tgcattatcg  | ttccctggga  | ggtgaaggca  | acttcaactg  | gaggttcatt | ttccccttcg  | 6000 |
| actacctgcc  | agctgagcaa  | gtctgtacca  | ttgccaaaga  | ggatgccttc | tggagctggg  | 6060 |
| acaagactga  | gagcaaaatc  | ccagcacgag  | tgggtgtcca  | gatctgggac | aatgacaagt  | 6120 |
| tctcctttga  | tgattttctg  | ggctccctgc  | agctcgatct  | caaccgcatg | cccaagccag  | 6180 |
| ccaagacagc  | caagaagtgc  | tccttggacc  | agctggatga  | tgctttccac | ccagaatggt  | 6240 |
| ttgtgtccct  | ttttgagcag  | aaaacagtga  | agggctgggt  | gccctgtgta | gcagaagagg  | 6300 |
| gtgagaagaa  | aatactggcg  | ggcaagctgg  | aaatgacctt  | ggagattgta | gcagagagtg  | 6360 |
| agcatgagga  | gcggcctgct  | ggccaggggc  | gggatgagcc  | caacatgaac | cctaagcttg  | 6420 |
| aggaccacaag | gcgccccgac  | acctccttcc  | tgtggtttac  | ctccccatac | aagaccatga  | 6480 |
| agttcatcct  | gtggcgcgct  | ttccggtggg  | ccatccatcc  | cttcatcatc | ctcttcatcc  | 6540 |
| tgctgctggt  | cctggccatc  | ttcatctacg  | ccttcccga   | ctatgctggc | atgaagctgg  | 6600 |
| tgaagccctt  | cagctgagga  | ctctcctgcc  | ctgtagaagg  | ggccgtgggg | ttccctccag  | 6660 |
| catgggactg  | gcctgcctcc  | tccgcccagc  | tcggcgagct  | cctccagacc | tcctaggcct  | 6720 |
| gattgtcctg  | ccagggtggg  | cagacagaca  | gatggaccgg  | cccacactcc | cagagttgct  | 6780 |
| aacatggagc  | tctgagatca  | ccccacttcc  | atcatttccct | tctcccccaa | cccaacgctt  | 6840 |

18/68

ttttggatca gctcagacat atttcagtat aaaacagttg gaaccacaaa aaaaaaaaaa 6900  
 aaaaaaaaaa aa 6912

&lt;210&gt; 14

&lt;211&gt; 6911

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 14

|             |             |             |             |            |             |      |
|-------------|-------------|-------------|-------------|------------|-------------|------|
| tcgaccgccc  | agccagggtgc | aaaatgccgt  | gtcattggga  | gactccgcag | ccggagcatt  | 60   |
| agattacagc  | tcgacgggagc | tcgggaaggg  | cggcgggggt  | ggaagatgag | cagaagcccc  | 120  |
| tggtctcgga  | acgcccggctg | acaagcgggg  | tgagcgcagg  | cggggcgggg | acccagccta  | 180  |
| gcccactgga  | gcagccgggg  | gtggcccgtt  | cccctttaag  | agcaactgct | ctaagccagg  | 240  |
| agccagagat  | tcgagccggc  | ctcgcccagc  | cagccctctc  | cagcgagggg | acccacaagc  | 300  |
| ggcgccctcg  | agcctccgac  | ctttccgagc  | cctctttgcg  | ccctgggcgc | acggggccct  | 360  |
| acacgcgcca  | agcatgctga  | gggtcttcat  | cctctatgcc  | gagaacgtcc | acacaccgga  | 420  |
| caccgacatc  | agcgatgcct  | actgctccgc  | gggtgttgca  | ggggtgaaga | agagaaccaa  | 480  |
| agtcatcaag  | aacagcgtga  | accctgtatg  | gaatgagggg  | tttgaatggg | acctcaaggg  | 540  |
| catccccctg  | gaccagggtc  | ctgagcttca  | tgtggtgggt  | aaagaccatg | agacgatggg  | 600  |
| gaggaacagg  | ttcctggggg  | aagccaaggt  | cccactccga  | gaggtcctcg | ccacccttag  | 660  |
| tctgtccgcc  | agcttcaatg  | ccccctgct   | ggacaccaag  | aagcagccca | caggggcctc  | 720  |
| gctggtcctg  | caggtgtcct  | acacaccgct  | gcctggagct  | gtgcccctgt | tcccgcctcc  | 780  |
| tactcctctg  | gagccctccc  | cgactctgcc  | tgacctggat  | gtagtggcag | acacaggagg  | 840  |
| agaggaagac  | acagaggacc  | agggactcac  | tggagatgag  | gcggagccat | tccctggatca | 900  |
| aagcggaggc  | ccgggggctc  | ccaccacccc  | aaggaaacta  | ccttcacgtc | ctccgccccca | 960  |
| ctaccccggg  | atcaaaaaga  | agcgaagtgc  | gcctacatct  | agaaagctgc | tgtcagacaa  | 1020 |
| accgcaggat  | ttccagatca  | gggtccaggt  | gatcgagggg  | cgccagctgc | cgggggtgaa  | 1080 |
| catcaagcct  | gtgggtcaagg | ttaccgctgc  | agggcagacc  | aagcggacgc | ggatccacaa  | 1140 |
| gggaaacagc  | ccactcttca  | atgagactct  | tttcttcaac  | ttgtttgact | ctcctgggga  | 1200 |
| gctggttgat  | gagcccatct  | ttatcacggt  | cgtagactct  | cgttctctca | ggacagatgc  | 1260 |
| tctcctcggg  | gagttccgga  | tggacgtggg  | caccatttac  | agagagcccc | ggcacgccta  | 1320 |
| tctcaggaag  | tggctgctgc  | tctcagaccc  | tgatgacttc  | tctgctgggg | ccagaggcta  | 1380 |
| cctgaaaaca  | agccttttgt  | tgtcggggcc  | tggggacgaa  | gcgcctctgg | agagaaaaga  | 1440 |
| cccctctgaa  | gacaaggagg  | acattgaaag  | caacctgctc  | cggcccacag | gcgtagccct  | 1500 |
| gcgaggagcc  | cacttctgcc  | tgaaggctct  | ccgggcccag  | gacttgccgc | agattgacga  | 1560 |
| tgccgtgatg  | gacaacgtga  | aacagatctt  | tggcttcgag  | agtaacaaga | agaacttggt  | 1620 |
| ggaccccttt  | gtggaggtca  | gctttgcggg  | gaaaatgctg  | tgcagcaaga | tcttgagaa   | 1680 |
| gacggccaac  | cctcagtggg  | accagaacat  | cacactgcct  | gccatgtttc | cctccatgtg  | 1740 |
| cgaaaaaatg  | aggattcgta  | tcatagactg  | ggaccgcctg  | actcacaatg | acatcgtggc  | 1800 |
| taccacctac  | ctgagtatgt  | cgaaaatctc  | tgccccctga  | ggagaaatag | aagaggagcc  | 1860 |
| tgcagggtgct | gtcaagcctt  | cgaaagcctc  | agacttggtg  | gactacctgg | gcttccctcc  | 1920 |
| cactttttggg | ccctgctaca  | tcaacctcta  | tggcagctcc  | agagagttca | caggcttccc  | 1980 |
| agacccttac  | acagagctca  | acacaggcaa  | gggggaaggt  | gtggcttata | gtggccggct  | 2040 |
| tctgctctcc  | ctggagacca  | agctggtgga  | gcacagtga   | cagaaggtgg | aggactgtcc  | 2100 |
| tgcggatgac  | atcctccggg  | tggagaagta  | ccttagagg   | cgcaagtact | ccctgtttgc  | 2160 |
| ggcctttctac | tcagccacca  | tgtgttagga  | tgtggatgat  | gccatccagt | ttgaggtcag  | 2220 |
| catcggaac   | tacgggaaca  | agttcgacat  | gacctgcctg  | ccgctggcct | ccaccactca  | 2280 |
| gtacagccgt  | gcagtctttg  | acgggtgcca  | ctactactac  | ctaccctggg | gtaacgtgaa  | 2340 |
| acctgtggtg  | gtgctgtcat  | cctactggga  | ggacatcagc  | catagaatcg | agactcagaa  | 2400 |
| ccagctgctt  | gggattgctg  | accggtgga   | agctggcctg  | gagcaggtcc | acctggccct  | 2460 |
| gaaggcgcag  | tgctccacgg  | aggacgtgga  | ctcgctggtg  | gctcagctga | cggatgagct  | 2520 |
| catcgaggc   | tgcagccagc  | ctctgggtga  | catccatgag  | acaccctctg | ccaccacact  | 2580 |
| ggaccagtac  | ctgtaccagc  | tgcgcaccca  | tcacctgagc  | caaatcactg | aggctgccct  | 2640 |
| ggccctgaag  | ctcgccaca   | gtgagctccc  | tgcagctctg  | gagcaggcgg | aggactggct  | 2700 |
| cctgcgtctg  | cgtgccctgg  | cagaggagcc  | ccagaacagc  | ctgccggaca | tcgtcatctg  | 2760 |
| gatgctgcag  | ggagacaagc  | gtgtggcata  | ccagcgggtg  | cccgccacc  | aagtccctct  | 2820 |
| ctcccgccgg  | ggtgccaaact | actgtggcaa  | gaattgtggg  | aagctacaga | caatctttct  | 2880 |
| gaaatatccg  | atggagaagg  | tgcctggcgc  | ccggtatgcca | gtgcagatac | gggtcaagct  | 2940 |
| tgtggtttggg | ctctctgtgg  | atgagaaggga | gttcaaccag  | tttgctgagg | ggaagtgtgc  | 3000 |
| gtcttttgct  | gaaacctatg  | agaacgagac  | taagtgtggc  | cttgttggga | actggggcac  | 3060 |
| aacgggcctc  | acctacccca  | agttttctga  | cgtcacgggc  | aagatcaagc | tacccaagga  | 3120 |
| cagcttccgc  | ccctcgcccg  | gctggacctg  | ggctggagat  | tggttcgtgt | gtccggagaa  | 3180 |
| gactctgctc  | catgacatgg  | acgcccgtca  | cctgagcttc  | gtggaagagg | tgtttgagaa  | 3240 |
| ccagaccccg  | cttcccgag   | gccagtggat  | ctacatgagt  | gacaactaca | ccgatgtgaa  | 3300 |
| cggggagaa   | gtgcttccca  | aggatgacat  | tgagtgccca  | ctgggctgga | agtggaaga   | 3360 |
| tgagggaatg  | tccacagacc  | tcaaccgggc  | tgtcgatgag  | caaggctggg | agtatagcat  | 3420 |
| caccatcccc  | ccggagcgga  | agccgaagca  | ctgggtccct  | gctgagaaga | tgtactacac  | 3480 |

19/68

|            |             |             |             |             |            |      |
|------------|-------------|-------------|-------------|-------------|------------|------|
| acaccgacgg | cggcgctggg  | tgcgcctgcg  | caggagggat  | ctcagccaaa  | tggaagcact | 3540 |
| gaaaaggcac | aggcaggcgg  | aggcggaggg  | cgaggggctg  | gagtagcgcct | ctcttttttg | 3600 |
| ctggaagttc | cacctcgagt  | accgcaagac  | agatgccttc  | cgccgcccgc  | gctggcgccg | 3660 |
| tcgcatggag | ccactggaga  | agacggggcc  | tgcagctgtg  | tttgcccttg  | agggggccct | 3720 |
| ggcgggcgtg | atggatgaca  | agagtgaaga  | ttccatgtcc  | gtctccacct  | tgagcttcgg | 3780 |
| tgtgaacaga | cccacgattt  | cctgcatatt  | cgactatggg  | aaccgctacc  | atctacgctg | 3840 |
| ctacatgtac | caggcccggg  | acctggctgc  | gatggacaag  | gactcttttt  | ctgatcccta | 3900 |
| tgccatcgtc | tccttcctgc  | accagagcca  | gaagacgggtg | gtggtgaaga  | acacccttaa | 3960 |
| ccccacctgg | gaccagacgc  | tcattcttcta | cgagatcgag  | atctttggcg  | agccggccac | 4020 |
| agttgctgag | caaccgcccc  | gcattgtggt  | ggagctgtac  | gaccatgaca  | cttatgggtg | 4080 |
| agacgagttt | atgggtcgct  | gcattctgtca | accgagctctg | gaacggatgc  | cacggctggc | 4140 |
| ctggttcccc | ctgagcaggg  | gcagccagcc  | gtcgggggag  | ctgctggcct  | cttttgagct | 4200 |
| catccagaga | gagaagccgg  | ccatccacca  | tattcctggt  | tttgaggtgc  | aggagacatc | 4260 |
| aaggatcctg | gatgagctgt  | aggacacaga  | cctgccctac  | ccaccacccc  | agagggagcg | 4320 |
| caacatctac | atggttcctc  | agaacatcaa  | gccagcgctc  | cagcgtaccg  | ccatcgagat | 4380 |
| cctggcatgg | ggcctgcgga  | acatgaagag  | ttaccagctg  | gccaacatct  | cctccccag  | 4440 |
| cctcgtggta | gagtggtggg  | gccagacggg  | gcagtcctgt  | gtcatcagga  | acctccggaa | 4500 |
| gaaccccaac | tttgacatct  | gcaccctctt  | catggaagtg  | atgctgcccc  | gggaggagct | 4560 |
| ctactgcccc | cccatcaccc  | tcaaggtcat  | cgataaaccg  | cagtttgccc  | gccggcctgt | 4620 |
| ggtggggccg | tgtaccatcc  | gctccctgga  | gagcttctctg | tgtgacctct  | actcggcgga | 4680 |
| gagtcctatc | ccacagggtg  | gccagagcga  | tgtgagccta  | ctcagtcctg  | gggaagacgt | 4740 |
| gctcatcgac | attgatgaca  | aggagcccc   | catccccatc  | caggaggaag  | agttcatcga | 4800 |
| ttggtggagc | aaattctttg  | cctccatagg  | ggagagggaa  | aagtgcggct  | cctacctgga | 4860 |
| gaaggatttt | gacaccctga  | aggtctatga  | cacacagctg  | gagaatgtgg  | agggagagac | 4920 |
| gggcctgtct | gacttttgtt  | acaccttcaa  | gctgtaccgg  | ggcaagacgc  | ccctcccaga | 4980 |
| agaagatcca | tctgtgattg  | gtgaatttaa  | gggcctcttc  | aaaattttatc | agggacccca | 5040 |
| agaccagcc  | atccccatgc  | ccccaaagaca | gttccaccag  | ctggccgccc  | aggaccccaa | 5100 |
| ggagtgcctg | gtccgtatct  | acattgtccg  | agcatttggc  | ctgcagccca  | aggaccccaa | 5160 |
| tggaagagtg | gatccttaca  | tcaagatctc  | catagggaa   | aatcagtgag  | gtgaccagga | 5220 |
| taactacatc | ccctgcacgc  | tggagcccgt  | atttggaag   | atgttcgagc  | tgacctgcac | 5280 |
| tctgcctctg | gagaaggacc  | taaagatcac  | tctctatgac  | tatgacctcc  | tctccaagga | 5340 |
| cgaaaagatc | ggtgagacgg  | tcgtcgacct  | ggagaacagg  | ctgctgtcca  | agtttggggc | 5400 |
| tcgctgtgga | ctcccacaga  | cctactgtgt  | ctctggaccg  | aaccagtggc  | gggagcagct | 5460 |
| ccgccccctc | cagctcctcc  | acctcttctg  | ccagcagcat  | agagtcaagg  | cacctgtgta | 5520 |
| ccggacagac | cgtgtaattg  | ttcaggataa  | agaatattcc  | attgaagaga  | tagaggctgg | 5580 |
| caggatccca | aacccacacc  | tgggcccagt  | ggaggagcgt  | ctggctctgc  | atgtgcttca | 5640 |
| gcagcagggc | ctgggtcccg  | agcacgtgga  | gtcacggccc  | ctctacagcc  | ccctgcagcc | 5700 |
| agacatcgag | caggggaagc  | tgcagatgtg  | ggtcgacctc  | tttccgaagg  | ccctggggcg | 5760 |
| gcctggacct | cccttcaaca  | tcacccacg   | gagagccaga  | aggtttttcc  | tgcgttgtat | 5820 |
| tatctggaat | accagagatg  | tgatcctgga  | tgacctgagc  | ctcacggggg  | agaagatgag | 5880 |
| cgacatttat | gtgaaagggt  | ggatgattgg  | ctttgaagaa  | cacaagcaaa  | agacagacgt | 5940 |
| gcattatcgt | tccttgaggag | gtgaaggcaa  | cttcaactgg  | aggttcattt  | ttcccttcga | 6000 |
| ctacctgcca | gctgagcaag  | tctgtaccat  | tgccaagaag  | gatgccttct  | ggaggctgga | 6060 |
| caagactgag | agcaaaatcc  | cagcacgagt  | ggtgttccag  | atctgggaca  | atgacaagtt | 6120 |
| ctcctttgat | gcttttctgg  | gctccctgca  | gctcgatctc  | aaccgcatgc  | ccaagccagc | 6180 |
| caagacagcc | aagaagtgtc  | ccttggaaca  | gctggatgat  | gctttccacc  | cagaatgggt | 6240 |
| tgtgtccctt | tttgagcaga  | aaacagtga   | gggctgggtg  | ccctgtgtag  | cagaagaggg | 6300 |
| tgagaagaaa | atactggcgg  | gcaagctgga  | aatgaccttg  | gagattgtag  | cagagagtg  | 6360 |
| gcatgaggag | cggcctgctg  | gccagggccg  | ggatgagccc  | aacatgaacc  | ctaagcttga | 6420 |
| ggacccaagg | cgccccgaca  | cctccttctc  | gtggtttacc  | tccccataca  | agaccatgaa | 6480 |
| gttcatcctg | tggcggcggt  | tccggtgggc  | catcatcctc  | ttcatcatcc  | tcttcatcct | 6540 |
| gctgctgttc | ctggccatct  | tcattctacgc | cttcccgaac  | tatgctgcca  | tgaagctggg | 6600 |
| gaagcccttc | agctgaggac  | tctcctgccc  | tgtagaaggg  | gccgtggggg  | cccctccagc | 6660 |
| atgggactgg | cctgcctcct  | ccgcccagct  | cggcgagctc  | ctccagacct  | cctaggcctg | 6720 |
| attgtcctgc | cagggtgggc  | agacagacag  | atggaccggc  | ccacactccc  | agagttgcta | 6780 |
| acatggagct | ctgagatcac  | cccacttcca  | tcatttctct  | ctcccccaac  | ccaacgcttt | 6840 |
| tttgatcag  | ctcagacata  | tttcagtata  | aaacagttgg  | aaccacaaaa  | aaaaaaaaaa | 6900 |
| aaaaaaaaaa | a           |             |             |             |            | 6911 |

&lt;210&gt; 15

&lt;211&gt; 6910

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 15

|            |            |            |            |            |            |     |
|------------|------------|------------|------------|------------|------------|-----|
| tcgaccgccc | agccagggtg | aaaatgccgt | gtcattggga | gactccgcag | ccggagcatt | 60  |
| agattacagc | tcgacggagc | tcgggaaggg | cggcgggggg | ggaagatgag | cagaagcccc | 120 |

|             |             |             |             |             |            |      |
|-------------|-------------|-------------|-------------|-------------|------------|------|
| tgttctcggg  | acgcccggctg | acaagcgggg  | tgagcgcag   | cggggcgggg  | accagccta  | 180  |
| gcccactgga  | gcagccgggg  | gtggcccgtt  | cccccttaag  | agcaactgct  | ctaagccagg | 240  |
| agccagagat  | tcgagccggc  | ctcgcccagc  | cagccctctc  | cagcgagggg  | accacaagc  | 300  |
| ggcgccctcg  | ccctcccagc  | ctttccgagc  | cctctttgcg  | ccctgggccc  | acggggccct | 360  |
| acacgcgcca  | agcatgctga  | gggtcttcat  | cctctatgcc  | gagaacgtcc  | acacacccga | 420  |
| caccgacatc  | agcgatgcct  | actgctccgc  | gggtgttgca  | ggggtgaaga  | agagaaccaa | 480  |
| agtcatacaag | aacagcgtga  | accctgtatg  | gaatgagggg  | tttgaatggg  | acctcaagg  | 540  |
| catccccctg  | gaccagggct  | ctgagcttca  | tgtggtggtc  | aaagaccatg  | agacgatggg | 600  |
| gaggaacagg  | ttcctggggg  | aagccaaggt  | cccactccga  | gaggtccctc  | ccacccttag | 660  |
| tctgtccgcc  | agcttcaatg  | cccccttgct  | ggacaccaag  | aagcagccca  | caggggccc  | 720  |
| gctggtcctg  | caggtgtcct  | acacacccgt  | gatcgagcgt  | gtgcccctgt  | tcccccccc  | 780  |
| tactcctctg  | gagccctccc  | cgactctgcc  | tgacctggat  | gtagtggcag  | acacaggagg | 840  |
| agaggaagac  | acagaggacc  | agggactcac  | tggagatgag  | gcggagccat  | tcctggatca | 900  |
| aagcggaggc  | ccgggggctc  | ccaccacccc  | aaggaaacta  | ccttcacgtc  | ctccgcccc  | 960  |
| ctaccccggg  | atcaaaaaga  | agcgaagtgc  | gcctacatct  | agaaagctgc  | tgtcagacaa | 1020 |
| accgcaggat  | ttccagatca  | gggtccaggt  | gatcgagggg  | cgccagctgc  | cgggggtgaa | 1080 |
| catcaagcct  | gtggtcaagg  | ttaccgctgc  | agggcagacc  | aagcggacgc  | ggatccacaa | 1140 |
| gggaaacagc  | ccactcttca  | atgagactct  | tttcttcaac  | ttgtttgact  | ctcctgggga | 1200 |
| gctgtttgat  | gagcccatct  | ttatcacggg  | ggtagactct  | cgttctctca  | ggacagatgc | 1260 |
| tctcctcggg  | gagttccgga  | tggacgtggg  | caccattttac | agagagcccc  | ggcacgccta | 1320 |
| tctcaggagg  | tggctgctgc  | tctcagaccc  | tgatgacttc  | tctgctgggg  | ccagaggcta | 1380 |
| cctgaaaaca  | agcctttgtg  | tgtgggggcc  | tggggacgaa  | gcgcctctgg  | agagaaaaga | 1440 |
| cccccttgaa  | gacaaggagg  | acattgaaag  | caacctgtct  | cggcccacag  | gcgtagccct | 1500 |
| gcgaggagcc  | cacttctgcc  | tgaaggctct  | ccgggcccag  | gacttgccgc  | agatggacga | 1560 |
| tgccgtgatg  | gacaacgtga  | aacagatctt  | tggcttcogag | agtaacaaga  | agaacttggt | 1620 |
| ggaccccttt  | gtgggtgctg  | gctttgctgg  | gaaaatgctg  | tgacgcaaga  | tcttgagaa  | 1680 |
| gacggccaac  | cctcagtggg  | accagaacat  | cacactgcct  | gccatgtttc  | cctccatgtg | 1740 |
| cgaaaaaatg  | aggattcgta  | tcatagactg  | ggaccgcctg  | actcacaatg  | acatcgtggc | 1800 |
| taccacctac  | ctgagtatgt  | cgaaaaatct  | tgcccctgga  | ggagaaaatg  | aagaggagcc | 1860 |
| tgcagggtgct | gtcaagcctt  | cgaaagcctc  | agacttggat  | gactacctgg  | gcttccctcc | 1920 |
| cacttttggtg | ccctgtctaca | tcaacctcta  | tggcagtcct  | agagagttca  | caggcttccc | 1980 |
| agacccttac  | acagagctca  | acacaggcaa  | gggggaaggt  | gtggcttatc  | gtggccggct | 2040 |
| tctgctctcc  | ctggagacca  | agctggtgga  | gcacagtga   | cagaagggtg  | aggaccttcc | 2100 |
| tgcggatgac  | atcctccggg  | tggagaagta  | ccttaggagg  | cgcaagtagt  | ccctgtttgc | 2160 |
| ggcctctctac | tcagccacca  | tgtcgcagga  | tgtggatgat  | gccatccagt  | ttgaggtcag | 2220 |
| catcgggaac  | tacgggaaca  | agttcgacat  | gactgcctcg  | ccgctggcct  | ccaccactca | 2280 |
| gtacagccgt  | gcagtctttg  | acgggtgcca  | ctactactac  | ctaccctggg  | gtaacgtgaa | 2340 |
| acctgtggtg  | gtgctgtcat  | cctactggga  | ggacatcagc  | catagaatcg  | agactcagaa | 2400 |
| ccagctgctt  | gggattgctg  | accggctgga  | agctggcctg  | gagcaggtcc  | acctggccct | 2460 |
| gaaggcgcag  | tgtctccagg  | aggacgtgga  | ctcgctgggt  | gctcagctga  | cggatgagct | 2520 |
| catcgcaggc  | tgagccagc   | ctctgggtga  | catccatgag  | acaccctctg  | ccaccacact | 2580 |
| ggaccagtag  | ctgtaccagc  | tgcgcaccca  | tcacctgagc  | caaatactctg | aggctgccct | 2640 |
| ggccctgaag  | ctcggccaca  | gtgagctccc  | tgcagctctg  | gagcaggcgg  | aggactggct | 2700 |
| cctgcgtctg  | cgtgccctgg  | cagaggagcc  | ccagaacagc  | ctgcccggaca | tcgtcatctg | 2760 |
| gatgtgcag   | ggagacaagc  | gtgtggcata  | ccagcgggtg  | cccgccacc   | aagtcctctt | 2820 |
| ctcccgcg    | gggtgccaact | actgtggcaa  | gaattgtggg  | aagctacaga  | caatctttct | 2880 |
| gaaatatccg  | atggagaagg  | tgcctggcgc  | ccggatgcca  | gtgcagatac  | gggtcaagct | 2940 |
| gtggtttggg  | ctctctgtgg  | atgagaagga  | gttcaaccag  | tttgctgagg  | ggaagctgtc | 3000 |
| tgtctttgct  | gaaacctatg  | agaacgagac  | taagttggcc  | cctgttggga  | actggggcac | 3060 |
| aacgggcctc  | acctacccca  | agttttctga  | cgtcacgggc  | aagatcaagc  | taccaagga  | 3120 |
| cagcttccgc  | ccctcgcccg  | gctggacctg  | ggctggagat  | tggttcgtgt  | gtccggagaa | 3180 |
| gactctgctc  | catgacatgg  | acgcccgtca  | cctgagcttc  | gtggaagagg  | tgtttgagaa | 3240 |
| ccagaccccg  | cttcccggag  | gccagtgat   | ctacatgagt  | gacaactaca  | ccgatgtgaa | 3300 |
| cggggagaag  | gtgcttccca  | aggatgacat  | tgagtgccta  | ctgggctgga  | agtgggaaga | 3360 |
| tgaggaaatg  | tccacagacc  | tcaaccgggc  | tgtcgatgag  | caaggctggg  | agtatagcat | 3420 |
| caccatcccc  | ccggagcgga  | agccgaagca  | ctgggtccct  | gctgagaaga  | tgtactacac | 3480 |
| acaccgacgg  | cggcgctggg  | tgcgcctgcg  | caggagggat  | ctcagccaaa  | tggagcact  | 3540 |
| gaaaaggcac  | aggcaggcgg  | aggcggaggg  | cgagggctgg  | gagtacgcct  | ctctttttgg | 3600 |
| tggaagtgc   | caactcgatc  | accgcaagac  | agatgccttc  | cgcgcgcgcc  | gctggcgccc | 3660 |
| tcgcatggag  | ccactggaga  | agacggggcc  | tgcagctgtg  | tttgcccttg  | agggggccct | 3720 |
| gggcggcgtg  | atggatgaca  | agagtgaaga  | ttccatgtcc  | gtctccacct  | tgagcttcgg | 3780 |
| tgtgaacaga  | cccacgattt  | cctgcatatt  | cgactatggg  | aaccgctacc  | atctacgctg | 3840 |
| ctacatgtac  | caggcccggg  | acctggctgc  | gatggacaag  | gactcttttt  | ctgatcccta | 3900 |
| tgccatcgtc  | tccttctctg  | accagagcca  | gaagacggtg  | gtggtgaaga  | acacccttaa | 3960 |
| ccccacctgg  | gaccagacgc  | tcactctcta  | cgagatcgag  | atctttggcg  | agccggccac | 4020 |
| agttgctgag  | caaccgcccc  | gcattgtggt  | ggagctgtac  | gaccatgaca  | cttatggtgc | 4080 |
| agacgagttt  | atgggtcgct  | gcactctgtca | accgagctct  | gaacggatgc  | cacggctggc | 4140 |

21/68

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| ctggttccca  | ctgacgaggg  | gcagccagcc  | gtcgggggag  | ctgctggcct  | cttttgagct  | 4200 |
| catccagaga  | gagaagccgg  | ccatccacca  | tattcctggg  | tttgaggtgc  | aggagacatc  | 4260 |
| aaggatcctg  | gatgagtcctg | aggacacaga  | cctgccctac  | ccaccacccc  | agagggagggc | 4320 |
| caacatctac  | atggttcctc  | agaacatcaa  | gccagcgctc  | cagcgtaccg  | ccatcgagat  | 4380 |
| cctggcatgg  | ggcctgcgga  | acatgaagag  | ttaccagctg  | gccaacatct  | cctccccag   | 4440 |
| cctcgtggta  | gagtggtggg  | gccagacggt  | gcagtcctgt  | gtcatcagga  | acctccggaa  | 4500 |
| gaaccccaac  | tttgacatct  | gcaccctctt  | catggaagtg  | atgctgcccc  | gggaggagct  | 4560 |
| ctactgcccc  | cccattaccg  | tcaagggtcat | cgataaccgc  | cagtttggcc  | gccggcctgt  | 4620 |
| ggtgggcccag | tgtaccatcc  | gctccctgga  | gagcttcctg  | tgtgaccctt  | actcggcgga  | 4680 |
| gagtcctatcc | ccacaggggtg | gcccagacga  | tgtgagccta  | ctcagtcctg  | gggaagacgt  | 4740 |
| gtcctatcgac | attgatgaca  | aggagccctt  | catccccatc  | caggaggaag  | agttcatcga  | 4800 |
| ttggtggagc  | aaattctttg  | cctccatagg  | ggagagggaa  | aagtgcggct  | cctacctgga  | 4860 |
| gaaggatttt  | gacaccctga  | aggtctatga  | cacacagctg  | gagaatgtgg  | aggcctttga  | 4920 |
| gggcctgtct  | gactttttgta | acaccttcaa  | gctgtaccgg  | ggcaagacgc  | aggaggagac  | 4980 |
| agaagatcca  | tctgtgattg  | gtgaatttaa  | gggcctcttc  | aaaatttatc  | ccctcccaga  | 5040 |
| agaccagcc   | atccccatgc  | cccccaagaca | gttccaccag  | ctggccgccc  | agggacccca  | 5100 |
| ggagtgcctg  | gtccgtatct  | acattgtccg  | agcatttggc  | ctgcagccca  | aggaccccaa  | 5160 |
| tggaaagtgt  | gatccttaca  | tcaagatctc  | catagggaag  | aaatcagtga  | gtgaccagga  | 5220 |
| taactacatc  | ccctgcacgc  | tggagcccgt  | atttggaag   | atgttcgagc  | tgacctgcac  | 5280 |
| tctgcctctg  | gagaaggacc  | tcaagatcac  | tctctatgac  | tatgacctcc  | tctccaagga  | 5340 |
| cgaaaagatc  | ggtgagacgg  | tcgtcgacct  | ggagaacagg  | ctgctgtcca  | agtttggggc  | 5400 |
| tcgctgtgga  | ctcccacaga  | cctactgtgt  | ctctggaccg  | aaccagtggc  | gggaccagct  | 5460 |
| ccgcccctcc  | cagctcctcc  | acctcttctg  | ccagcagcat  | agagtcaagg  | cacctgtgta  | 5520 |
| ccggacagac  | cgtgtaatgt  | ttcaggataa  | agaatatctc  | attgaagaga  | tagaggcttg  | 5580 |
| caggatccca  | aacccacacc  | tgggccaggt  | ggaggagcgt  | ctggctctgc  | atgtgttca   | 5640 |
| gcagcagggc  | ctggtcccgg  | agcacgtgga  | gtcacggccc  | ctctacagcc  | ccctgcagcc  | 5700 |
| agacatcgag  | caggggaagc  | tgcagatgtg  | ggtcgacctc  | tttccgaagg  | ccctggggcg  | 5760 |
| gcctggacct  | cccttcaaca  | tcaccccacg  | gagagccaga  | agggttttcc  | tgcgttgtat  | 5820 |
| tatctggaat  | accagagatg  | tgatcctgga  | tgacctgagc  | ctcacggggg  | agaagatgag  | 5880 |
| cgacatttat  | gtgaaagggt  | ggatgattgg  | ctttgaagaa  | cacaagcaaa  | agacagacgt  | 5940 |
| gcattatcgt  | tccctgggag  | gtgaagcaac  | ttcaactgga  | ggttcatttt  | ccccttcgac  | 6000 |
| tacctgccag  | ctgagcaagt  | ctgtaccatt  | gccaagaagg  | atgccttctg  | gaggctggac  | 6060 |
| aagactgaga  | gcaaaatccc  | agcacgagtg  | gtgttccaga  | tctgggacaa  | tgacaagttc  | 6120 |
| tcctttgatg  | atthttctggg | ctccctgcag  | ctcgatctca  | accgcatgcc  | caagccagcc  | 6180 |
| aagacagcca  | agaagtgtct  | cttgaccag   | ctggatgatg  | ctttccaccc  | agaatggttt  | 6240 |
| gtgtcccttt  | ttgagcagaa  | aacagtgaag  | ggctggtggc  | cctgtgtagc  | agaagaggggt | 6300 |
| gagaagaaaa  | tactggcggg  | caagctggaa  | atgaccttgg  | agattgtagc  | agagagttag  | 6360 |
| catgaggagc  | ggcctgctgg  | ccagggccgg  | gatgagccca  | acatgaaccc  | taagcttgag  | 6420 |
| gacccaaggc  | gccccgacac  | ctccttctctg | tgggtttact  | ccccatacaa  | gacctgaag   | 6480 |
| ttcatcctgt  | ggcggcgttt  | ccggtgggcc  | atcatcctct  | tcacatcct   | cttcatcctg  | 6540 |
| ctgctgttcc  | tggccatctt  | catctacgcc  | ttcccgaaact | atgctgccat  | gaagctgggtg | 6600 |
| aagcccttca  | gctgaggact  | ctcctgccct  | gtagaagggg  | ccgtgggggtc | ccctccagca  | 6660 |
| tgggactggc  | ctgcctcctc  | cgcccagctc  | ggcgagctcc  | tccagacctc  | ctaggcctga  | 6720 |
| ttgtcctgcc  | aggggtgggca | gacagacaga  | tggaccggcc  | cacactccca  | gagttgtgtaa | 6780 |
| catggagctc  | tgagatcacc  | ccacttccat  | catcttcttc  | tcccccaacc  | caacgctttt  | 6840 |
| ttggatcagc  | tcagacatat  | ttcagtataa  | aacagttgga  | accacaaaaa  | aaaaaaaaaa  | 6900 |
| aaaaaaaaaa  |             |             |             |             |             | 6910 |

<210> 16  
 <211> 6911  
 <212> DNA  
 <213> Homo sapiens

|             |             |            |             |            |            |     |
|-------------|-------------|------------|-------------|------------|------------|-----|
| <400> 16    |             |            |             |            |            |     |
| tgcaccgccc  | agccaggtgc  | aaaatgccgt | gtcattggga  | gactccgcag | ccggagcatt | 60  |
| agattacagc  | tcgacggagc  | tcgggaagg  | cggcggggg   | ggaagatgag | cagaagcccc | 120 |
| tggttctcga  | acgcccggctg | acaagcgggg | tgagcgcagg  | cggggcgggg | acccagccta | 180 |
| gcccactgga  | gcagccgggg  | gtggcccgtt | cccctttaag  | agcaactgct | ctaagccagg | 240 |
| agccagagat  | tcgagccggc  | ctcgcccagc | cagccctctc  | cagcgagggg | acccacaagc | 300 |
| ggcgccctcg  | ccctcccagc  | ctttccgagc | cctctttgcg  | ccctggggcg | acggggccct | 360 |
| acacgcgcca  | agcatgctga  | gggtcttcat | cctctatgcc  | gagaacgtcc | acacacccga | 420 |
| caccgacatc  | agcgaatgct  | actgctccgc | ggtgtttgca  | ggggtgaaga | agagaaccaa | 480 |
| agtcatcaag  | aacagcgtga  | accctgtatg | gaatgagggg  | tttgaatggg | acctcaagg  | 540 |
| catccccctg  | gaccagggtc  | ctgagcttca | tgtggtgggtc | aaagaccatg | agacgatggg | 600 |
| gaggaacagg  | ttcctggggg  | aagccaaagt | cccactccga  | gaggtcctcg | ccacccctag | 660 |
| tctgtccggc  | agcttcaatg  | ccccctgct  | ggacaccaag  | aagcagccca | caggggcctc | 720 |
| gctgggtcctg | caggtgtcct  | acacaccgct | gcctggagct  | gtgccctgt  | ttccgcccc  | 780 |

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| tactcctctg  | gagccctccc  | cgactctgcc  | tgacctggat  | gtagtggcag  | acacaggagg  | 840  |
| agaggaagac  | acagaggacc  | agggactcac  | tggagatgag  | gcggagccat  | tcctggatca  | 900  |
| aagcggaggc  | ccgggggctc  | ccaccacccc  | aaggaaacta  | ccttcacgtc  | ctccgccccca | 960  |
| ctaccccggg  | atcaaaaagaa | agcgaagtgc  | gcctacatct  | agaaagctgc  | tgtcagacaa  | 1020 |
| accgcaggat  | ttccagatca  | gggtccaggt  | gatcgagggg  | cgccagctgc  | cggggggtgaa | 1080 |
| catcaagcct  | gtggtcaagg  | ttaccgctgc  | agggcagacc  | aagcggacgc  | ggatccacaa  | 1140 |
| gggaaacagc  | ccactcttca  | atgagactct  | tttcttcaac  | ttgtttgact  | ctcctgggga  | 1200 |
| gctgtttgat  | gagcccatct  | ttatcacggt  | ggtagactct  | cgttctctca  | ggacagatgc  | 1260 |
| tctcctcggg  | gagttccgga  | tggacgtggg  | caccatttac  | agagagcccc  | ggcacgccta  | 1320 |
| tctcaggaag  | tggtctgtgc  | tctcagaccc  | tgatgacttc  | tctgctgggg  | ccagaggcta  | 1380 |
| cctgaaaaca  | agcctttgtg  | tgttggggcc  | tggggacgaa  | gcgcctctgg  | agagaaaaga  | 1440 |
| ccccctgaa   | gacaaggagg  | acattgaaag  | caacctgtct  | cggccccacag | gcgtagccct  | 1500 |
| gcgaggagcc  | cacttctgcc  | tgaaggtctt  | ccgggcccag  | gacttgccgc  | agatggacga  | 1560 |
| tgccgtgatg  | gacaacgtga  | aacagatctt  | tggcttcgag  | agtaacaaga  | agaacttggt  | 1620 |
| ggacccccctt | gtggaggtca  | gctttgcggg  | gaaaatgctg  | tgcagcaaga  | tcttggagaa  | 1680 |
| gacggccaac  | cctcagtgga  | accagaacat  | cacactgcct  | gccatgtttc  | cctccatgtg  | 1740 |
| cgaaaaaatg  | aggattcgta  | tcatagactg  | ggaccgcctg  | actcacaatg  | acatcgtggc  | 1800 |
| taccacctac  | ctgagtatgt  | cgaaaatctc  | tgccccctgga | ggagaaatag  | aagaggagcc  | 1860 |
| tgcaggtgct  | gtcaagcctt  | cgaaagcctc  | agacttggtg  | gactacctgg  | gcttctctcc  | 1920 |
| cacttttggg  | ccctgctaca  | tcaacctcta  | tggcagtcct  | agagagtcca  | caggcttccc  | 1980 |
| agacccctac  | acagagctca  | acacaggcaa  | gggggaaggt  | gtggcttatc  | gtggccggct  | 2040 |
| tctgctctcc  | ctggagacca  | agctggtgga  | gcacagtga   | cagaaggtgg  | aggaccttcc  | 2100 |
| tgcggatgac  | atcctccggg  | tggagaagta  | ccttaggagg  | cgcaagtact  | ccctgtttgc  | 2160 |
| ggccttctac  | tcagccacca  | tgtgcagga   | tgtggatgat  | gccatccagt  | ttgaggtcag  | 2220 |
| catcgggaac  | tacgggaaca  | agttcgacat  | gacctgcctg  | ccgttggcct  | ccaccactca  | 2280 |
| gtacagccgt  | cgagtctttg  | acgggtgcca  | ctactactac  | ctacctgggg  | gtaacgtgaa  | 2340 |
| acctgtggtg  | gtgctgtcat  | cctactggga  | ggacatcagc  | catagaatcg  | agactcagaa  | 2400 |
| ccagctgctt  | gggattgtctg | accggctgga  | agctggcctg  | gagcagggtcc | acctggccct  | 2460 |
| gaaggcgcag  | tgctccacgg  | aggacgtgga  | ctcgctgggtg | gctcagctga  | cggatgagct  | 2520 |
| catcgcaggc  | tgcagccagc  | ctctgggtga  | catccatgag  | acacctctctg | ccaccacact  | 2580 |
| ggaccagtag  | ctgtaccagc  | tgcgcaccga  | tcacctgagc  | caaatcactg  | aggctgcctt  | 2640 |
| ggccctgaag  | ctcggccaca  | gtgagctccc  | tgcagctctg  | gagcaggcgg  | aggactggct  | 2700 |
| cctgcgtctg  | cgtgccctgg  | cagaggagcc  | ccagaacagc  | ctgcccggaca | tcgtcatctg  | 2760 |
| gatgctgcag  | ggagacaagc  | gtgtggcata  | ccagcgggtg  | cccgccacc   | aagtctcttt  | 2820 |
| ctcccggcgg  | ggtgccaaact | actgtggcaa  | gaattgtggg  | aagctacaga  | caattcttct  | 2880 |
| gaaatatccg  | atggagaagg  | tgcctggcgc  | ccggatgcca  | gtgcagatac  | gggtcaagct  | 2940 |
| gtggtttggg  | ctctctgtgg  | atgagaagga  | gttcaaccag  | tttgcgtgagg | ggaagctgtc  | 3000 |
| tgtctttgct  | gaaacctatg  | agaacgagac  | taagtgtggc  | cctgttggga  | actggggcac  | 3060 |
| aacgggcctc  | acctaccccc  | agttttctga  | cgtcacgggc  | aagatcaagc  | taccacaagg  | 3120 |
| cagctttccg  | ccctgcgcg   | gctggacctg  | ggctggagat  | tggttcgtgt  | gtccggagaa  | 3180 |
| gactctgctc  | catgacatgg  | acgccgttca  | cctgagcttc  | gtggaagagg  | tgtttgagaa  | 3240 |
| ccagaccggg  | cttcccggag  | gccagtggat  | ctacatgagt  | gacaactaca  | ccgatgtgaa  | 3300 |
| cggggagaaag | gtgcttcccc  | aggatgacat  | tgagtgccta  | ctgggctgga  | agtgggaaga  | 3360 |
| tgaggaatgg  | tccacagacc  | tcaaccgggc  | tgtcgatgag  | caaggctggg  | agtatagcat  | 3420 |
| caccatcccc  | cgggagcgga  | agccgaagca  | ctgggtccct  | gctgagaaga  | tgtactacac  | 3480 |
| acaccgacgg  | cggcgctggg  | tgcgcctgcg  | caggagggat  | ctcagccaaa  | tggaaagcact | 3540 |
| gaaaaggcac  | aggcaggcgg  | aggcggaggg  | cgagggtgg   | gagtacgcct  | ctctttttgg  | 3600 |
| ctggaagttc  | cacctcgagt  | accgcaagac  | agatgccttc  | cgcgcgccgc  | gctggcgccg  | 3660 |
| tcgcatggag  | ccactggaga  | agacggggcc  | tgcagctgtg  | tttgcccttg  | agggggccct  | 3720 |
| gggcggcgtg  | atggatgaca  | agagtgaaga  | ttccatgtcc  | gtctccacct  | tgagcttcgg  | 3780 |
| tgtgaacaga  | cccacgattt  | cctgcatatt  | cgactatggg  | aaccgctacc  | atctacgctg  | 3840 |
| ctacatgtac  | caggcccggg  | acctggctgc  | gatggacaag  | gactcttttt  | ctgatcccta  | 3900 |
| tgccatcgte  | tccttccctg  | accagagcca  | gaagacgggtg | gtggtgaaga  | acacccttaa  | 3960 |
| ccccacctgg  | gaccagacgc  | tcatcttcta  | cgagatcgag  | atcttttggcg | agccggccac  | 4020 |
| agttgctgag  | caaccgcccc  | gcattgtggt  | ggagctgtac  | gaccatgaca  | cttatgggtgc | 4080 |
| agacgagttt  | atgggtcgct  | gcattgttca  | accgagctctg | gaacggatgc  | cacggctggc  | 4140 |
| ctggttccca  | ctgacgaggg  | gcagccagcc  | gtcgggggag  | ctgctggcct  | cttttgagct  | 4200 |
| catccagaga  | gagaagccgg  | ccatccacca  | tattcctggt  | tttgaggtgc  | aggagacatc  | 4260 |
| aaggatccctg | gatgagctctg | aggacacaga  | cctgccctac  | ccaccacccc  | agaggagggc  | 4320 |
| caacatctac  | atggttctctc | agaacatcaa  | gccagcgctc  | cagcgtaccg  | ccatcgagat  | 4380 |
| cctggcatgg  | ggcctgcgga  | acatgaagag  | ttaccagctg  | gccaacatct  | cctccccag   | 4440 |
| cctcgtggtg  | gagtggtggg  | gccagacggt  | gcagtcctgt  | gtcatcagga  | acctccggaa  | 4500 |
| gaaccccaac  | tttgacatct  | gcacctctt   | catggaagtg  | atgctgcccc  | gggaggagct  | 4560 |
| ctactgcccc  | cccatcaccg  | tcaaggtcat  | cgataaccgc  | cagtttggcc  | gccggcgtgt  | 4620 |
| ggtgggcccag | tgtaccatcc  | gctccctgga  | gagcttctctg | tgtgacccct  | actcggcgga  | 4680 |
| gagtcctatcc | ccacagggtg  | gcccagacga  | tgtgagccta  | ctcagtcctg  | gggaagacgt  | 4740 |
| gctcatcgac  | attgatgaca  | aggagccccct | catccccatc  | caggaggaag  | agttcatcga  | 4800 |



23/68

|             |             |             |             |            |            |      |
|-------------|-------------|-------------|-------------|------------|------------|------|
| ttggtggagc  | aaattctttg  | cctccatag   | ggagagggaa  | aagtgcggct | cctacctgga | 4860 |
| gaaggatttt  | gacacctga   | aggtctatga  | cacacagctg  | gagaatgtgg | aggcctttga | 4920 |
| gggcctgtct  | gacttttgta  | acaccttcaa  | gctgtaccgg  | ggcaagacgc | aggaggagac | 4980 |
| agaagatcca  | tctgtgattg  | gtgaatttaa  | gggcctcttc  | aaaatttatc | ccctcccaga | 5040 |
| agaccagcc   | atccccatgc  | ccccaaagaca | gttccaccag  | ctggccgccc | agggacccca | 5100 |
| ggagtgttg   | gtccgtatct  | acattgtccg  | agcatttggc  | ctgcagccca | aggaccccaa | 5160 |
| tggaaaagtgt | gatccttaca  | tcaagatctc  | catagggaaag | aaatcagtga | gtgaccagga | 5220 |
| taactacatc  | ccctgcacgc  | tggagcccg   | atgttcgagc  | atgttcgagc | tgacctgcac | 5280 |
| tctgcctctg  | gagaaggacc  | taaagatcac  | tctctatgac  | tatgacctcc | tctccaagga | 5340 |
| cgaaaagatc  | ggtgagacgg  | tcgtcgacct  | ggagaacagg  | ctgctgtcca | agtttggggc | 5400 |
| tcgctgtgga  | ctcccacaga  | cctactgtgt  | ctctggaccg  | aaccagtggc | gggaccagct | 5460 |
| cgccccctcc  | cagctcctcc  | acctcttctg  | ccagcagcat  | agagtcaagg | cacctgtgta | 5520 |
| ccggacagac  | cgtgtaattgt | ttcaggataa  | agaatattcc  | attgaagaga | tagaggctgg | 5580 |
| caggatccca  | aaccacacac  | tgggcccagt  | ggaggagcgt  | ctggctctgc | atgtgcttca | 5640 |
| gcagcagggc  | ctggtcccgc  | agcacgtgga  | gtcacggccc  | ctctacagcc | ccctgcagcc | 5700 |
| agacatcgag  | caggggaagc  | tgcatgtgtg  | ggctcgacct  | tttccgaagg | ccctggggcg | 5760 |
| gcctggacct  | cccttcaaca  | tcacccacag  | gagagccaga  | aggtttttcc | tgctgtgtat | 5820 |
| tatctggaat  | accagagatg  | tgatcctgga  | tgacctgagc  | ctcacgggg  | agaagatgag | 5880 |
| cgacatttat  | gtgaaaaggtt | ggatgattgg  | ctttgaagaa  | cacaagcaaa | agacagacgt | 5940 |
| gcattatcgt  | tccctgggag  | gtgaaggcaa  | cttcaactgg  | aggttcattt | tcccttcga  | 6000 |
| ctacctgcca  | gctgagcaag  | tctgtacct   | tgccaagaag  | gatgccttct | ggaggctgga | 6060 |
| caagactgag  | agcaaaatcc  | cagcacgagt  | ggtgttccag  | atctgggaca | atgacaagtt | 6120 |
| ctcctttgat  | gattttctgg  | gctccctgca  | gctcgatctc  | aaccgcagtc | ccaagccagc | 6180 |
| caagacagcc  | aagaagtgtct | ccttggacca  | gctggatgat  | gctttccacc | cagaatgggt | 6240 |
| tgtgtccctt  | tttgagcaga  | aaacagtga   | gggctgggtg  | ccctgtgtag | cagaagagg  | 6300 |
| tgagaagaaa  | atactggcg   | gcaagctgga  | aatgaccttg  | gagattgtag | cagagagtga | 6360 |
| gcatgaggag  | cgccctgctg  | gccagggccg  | ggatgagccc  | aacatgaacc | ctaagcttga | 6420 |
| ggacccaagg  | cgccccgaca  | cctccttcc   | gtgggttacc  | tccccataca | agaccatgaa | 6480 |
| gttcatcctg  | tgccggcggt  | tccggtgggc  | catcatctc   | ttcatcatcc | tcttcatcct | 6540 |
| gctgctgttc  | ctggccatct  | tcactacgc   | cttcccgaac  | tatgctgcca | tgaagctggt | 6600 |
| gaagcccttc  | agctgaggac  | tctcctggcc  | tgtagaagg   | gccgtgggg  | cccctccagc | 6660 |
| atgggactgg  | cctgcctcct  | ccgcccagct  | cgccgagctc  | ctccagacct | cctaggcctg | 6720 |
| attgtcctgc  | cagggtgggc  | agacagacag  | atggaccggc  | ccacactccc | agagttgcta | 6780 |
| acatggagct  | ctgagatcac  | cccacttcca  | tcatttccct  | ctcccccaac | ccaacgcctt | 6840 |
| tttggatcag  | ctcagacata  | tttcagtata  | aaacagtgg   | aaccacaaaa | aaaaaaaaa  | 6900 |
| aaaaaaaaa   | a           |             |             |            |            | 6911 |

&lt;210&gt; 17

&lt;211&gt; 6911

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 17

|             |             |            |            |            |             |      |
|-------------|-------------|------------|------------|------------|-------------|------|
| tcgaccgccc  | agccaggtgc  | aaaatgccgt | gtcattggga | gactccgcag | ccggagcatt  | 60   |
| agattacagc  | tcgacggagc  | tcgggaaggt | cgccgggggt | ggaagatgag | cagaagcccc  | 120  |
| tgttctcggg  | acgccggctg  | acaagcgggg | tgagcgcagg | cggggcgggg | accagcccta  | 180  |
| gcccactgga  | gcagccgggg  | gtggcccggt | cccctttaag | agcaactgct | ctaagccagg  | 240  |
| agccagagat  | tcgagccggc  | ctcgcccagc | cagccctctc | cagcgagggg | accacaagc   | 300  |
| ggcgctcgg   | ccctcccagc  | ctttccgagc | cctctttgcg | ccctgggcgc | acggggccct  | 360  |
| acacgcgcca  | agcatgctga  | gggtcttcat | cctctatgcc | gagaacgtcc | acacaccgga  | 420  |
| caccgacatc  | agcgatgcct  | actgctccgc | ggtgtttgca | ggggtgaaga | agagaaccaa  | 480  |
| agtcatcaag  | aacagcgtga  | accctgtatg | gaatgaggga | tttgaatggg | acctcaagg   | 540  |
| catccccctg  | gaccagggct  | ctgagcttca | tgtgggtggc | aaagaccatg | agacgatggg  | 600  |
| gaggaacagg  | ttcctggggg  | aagccaaggt | cccactccga | gaggtcctcg | ccacccttag  | 660  |
| tctgtccgccc | agcttcaatg  | ccccctgct  | ggacaccaag | aagcagccca | caggggcctc  | 720  |
| gctggctcctg | caggtgtcct  | acacaccgct | gcctggagct | gtgcccctgt | tcccgcccc   | 780  |
| tactcctctg  | gagccctccc  | cgactctgcc | tgacctggat | gtagtggcag | acacaggagg  | 840  |
| agaggaagac  | acagaggacc  | agggactcac | tggagatgag | gcggagccat | tcctggatca  | 900  |
| aagcggaggc  | ccgggggctc  | ccaccacccc | aaggaaacta | ccttcacgtc | ctccgcccc   | 960  |
| ctaccccggg  | atcaaaaagaa | agcgaagtgc | gcctacatct | agaaagctgc | tgtcagacaa  | 1020 |
| accgcaggat  | ttccagatca  | gggtccaggt | gatcgagggg | cgccagctgc | cggggggtgaa | 1080 |
| catcaagcct  | gtgggtcaagg | ttaccgctgc | agggcagacc | aagcggacgc | ggatccacaa  | 1140 |
| gggaaacagc  | ccactcttca  | atgagactct | tttcttcaac | ttgtttgact | ctcctgggga  | 1200 |
| gctgtttgat  | gagcccatct  | ttatcacggt | ggtagactct | cgttctctca | ggacagatgc  | 1260 |
| tctcctcggg  | gagttccgga  | tggacgtggg | caccatttac | agagagcccc | ggcacgccta  | 1320 |
| tctcaggaag  | tggctgctgc  | tctcagaccc | tgatgacttc | tctgctgggg | ccagaggcta  | 1380 |
| cctgaaaaca  | agcctttgtg  | tgctggggcc | tggggacgaa | gcgcctctgg | agagaaaaga  | 1440 |

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| ccccctctgaa | gacaaggagg  | acattgaaag  | caacctgctc  | cgggccacag  | gcgtagccct  | 1500 |
| gcgaggagcc  | cacttctgcc  | tgaaggtctt  | ccggggccgag | gacttgccgc  | agatggacga  | 1560 |
| tgccgtgatg  | gacaacgtga  | aacagatctt  | tggcttcgag  | agtaacaaga  | agaacttggt  | 1620 |
| ggaccctctt  | gtggaggtca  | gctttgcggg  | gaaaatgctg  | tgacgcaaga  | tcttggagaa  | 1680 |
| gacggccaac  | cctcagtgga  | accagaacat  | cacactgcct  | gcatgtttc   | cctccatgtg  | 1740 |
| cgaaaaaatg  | aggattcgta  | tcatagactg  | ggaccgcctg  | actcacaatg  | acatcgaggc  | 1800 |
| taccacctac  | ctgagtatgt  | cgaaaaatctc | tgcccctgga  | ggagaaatag  | aagaggagcc  | 1860 |
| tgacggtgct  | gtcaagcctt  | cgaaagcctc  | agacttggat  | gactacctgg  | gcttccctccc | 1920 |
| cacttttggg  | ccctgctaca  | tcaacctcta  | tggcagtcct  | agagagttca  | caggcttccc  | 1980 |
| agacccttac  | acagagctca  | acacaggcaa  | gggggaaggt  | gtggcttata  | gtggccggct  | 2040 |
| tctgctctcc  | ctggagacca  | agctgggtga  | gcacagtga   | cagaaggtgg  | aggaccttcc  | 2100 |
| tgccgatgac  | atcctccggg  | tggagaagta  | ccttaggagg  | cgcaagtact  | ccctgttttg  | 2160 |
| ggccttctac  | tcagccacca  | tgctgcagga  | tgtggatgat  | gccatccagt  | ttgaggtcag  | 2220 |
| catcggaac   | tacgggaaca  | agttcgacat  | gacctgcctg  | ccgctggcct  | ccaccactca  | 2280 |
| gtacagccgt  | gcagtctttg  | acgggtgcca  | ctactactac  | ctaccctggg  | gtaacgtgaa  | 2340 |
| acctgtggtg  | gtgctgtcat  | cctactggga  | ggacatcagc  | catagaatcg  | agactcagaa  | 2400 |
| ccagctgctt  | gggattgctg  | accggctgga  | agctggcctg  | gagcaggtcc  | acctggccct  | 2460 |
| gaaggcgag   | tgctccacgg  | aggacgtgga  | ctcgctgggt  | gctcagctga  | cggatgagct  | 2520 |
| catcgaggg   | tgacgacagc  | ctctgggtga  | catccatgag  | acaccctctg  | ccaccacact  | 2580 |
| ggaccagtac  | ctgtaccagc  | tgcgcaccca  | tcacctgagc  | caaatactctg | aggctgccct  | 2640 |
| ggccctgaag  | ctcgccaca   | gtgagctccc  | tgagctctg   | gagcaggcgg  | aggactggct  | 2700 |
| cctgcgtctg  | cgctgccctg  | cagaggagcc  | ccagaacagc  | ctgccggaca  | tcgtcatctg  | 2760 |
| gatgctgcag  | ggagacaagc  | gtgtggcata  | ccagcgggtg  | cccgcccacc  | aagtcctctt  | 2820 |
| ctcccggcgg  | ggtgccaaact | actgtggcaa  | gaattgtggg  | aagctacaga  | caatctttct  | 2880 |
| gaaatatccg  | atggagaagg  | tgccctggcg  | ccggatgccca | gtgcagatac  | gggtcaagct  | 2940 |
| gtggtttggg  | ctctctgtgg  | atgagaagca  | ggtcaaccag  | tttgcgtgag  | ggagagctgt  | 3000 |
| tgtctttgct  | gaaacctatg  | agaacgagac  | taagtgtggc  | cttgtttggg  | actggggcac  | 3060 |
| aacgggcctc  | acctacccca  | agttttctga  | cgtcacgggc  | aagatcaagc  | taccaagga   | 3120 |
| cagcttccgc  | ccctcgcccg  | gctggacctg  | ggctggagat  | tggttcgtgt  | gtccggagaa  | 3180 |
| gactctgctc  | catgacatgg  | acgccgggtc  | cctgagcttc  | gtggaagagg  | tggttgagaa  | 3240 |
| ccagaccctg  | cttcccggag  | gccagtggat  | ctacatgagt  | gacaactaca  | ccgatgtgaa  | 3300 |
| cggggagaag  | gtgcttccca  | aggatgacat  | tgagtgccca  | ctgggctgga  | agtgggaaga  | 3360 |
| tgaggaatgg  | tccacagacc  | tcaaccgggc  | tgtcgatgag  | caaggctggg  | agtatagcat  | 3420 |
| caccatcccc  | ccggagcgga  | agccgaagca  | ctgggtccct  | gctgagaaga  | tgtactacac  | 3480 |
| acaccgacgg  | cggcgctggg  | tgccgctgcg  | caggagggat  | ctcagccaaa  | tggaagcact  | 3540 |
| gaaaaggcac  | aggcggaggg  | aggcggaggg  | cgagggtctg  | gagtacgcct  | ctcttttttg  | 3600 |
| ctggaagttc  | cacctcgagt  | accgcaagac  | agatgccttc  | cgccgcccgc  | gctggcgccg  | 3660 |
| tcgcatggag  | ccactggaga  | agacggggcc  | tgacgtgtg   | tttgcccttg  | agggggccct  | 3720 |
| gggcggcggt  | atggatgaca  | agagtgaaga  | ttccatgtcc  | gtctccacct  | tgagcttcgg  | 3780 |
| tgtgaacaga  | cccacgattt  | cctgcatatt  | cgactatggg  | aaccgctacc  | atctacgtcg  | 3840 |
| ctacatgtac  | caggcccggg  | acctggctgc  | gatggacaag  | gactcttttt  | ctgatcccta  | 3900 |
| tgccatcgtc  | tccttccctg  | accagagcca  | gaagacggtg  | gtggtgaaga  | acacccttaa  | 3960 |
| ccccacctgg  | gaccagacgc  | tcatcttcta  | cgagatcgag  | atctttggcg  | agccggccac  | 4020 |
| agttgctgag  | caaccgcccc  | gcatttgtgt  | ggagctgtac  | gaccatgaca  | cttatggtgc  | 4080 |
| agacgagttt  | atgggtcgct  | gcacgtgtca  | accagatctg  | gaacggatgc  | cacggctggc  | 4140 |
| ctgggttccca | ctgacgaggg  | gcagccagcc  | gtcgggggag  | ctgctggcct  | cttttgagct  | 4200 |
| catccagaga  | gagaagccgg  | ccatccacca  | tattcctggg  | tttgaggtgc  | aggagacatc  | 4260 |
| aaggatcctg  | gatgagtctg  | aggacacaga  | cctgccctac  | ccaccacccc  | agagggaggc  | 4320 |
| caacatctac  | atggttccct  | agaacatcaa  | gccagcgctc  | cagcgtaccg  | ccatcgagat  | 4380 |
| cctggcatgg  | ggcctgcgga  | acatgaagag  | ttaccagctg  | gccaacatct  | cctccccag   | 4440 |
| ctcgtggta   | gagtgtgggg  | gccagacggt  | gcagtcctgt  | gtcatcagga  | acctccggaa  | 4500 |
| gaaccccaac  | tttgacatct  | gcaccctctt  | catggaagtg  | atgctgccca  | gggaggagct  | 4560 |
| ctactgcccc  | cccatacccg  | tcaaggtcat  | cgataaaccg  | cagtttggcc  | gccggcctgt  | 4620 |
| ggtgggcccag | tgtaccatcc  | gctccctgga  | gagcttctct  | tgtgaccctt  | actcggcgga  | 4680 |
| gagtcctatcc | ccacagggtg  | gccagacga   | tgtgagccta  | ctcagtcctg  | gggaagacgt  | 4740 |
| gctcatcgac  | attgatgaca  | aggagccctt  | catccccatc  | caggaggaag  | agttcatcga  | 4800 |
| ttggtggagc  | aaattctttg  | cctccatagg  | ggagagggaa  | aagtgcggct  | cctacctgga  | 4860 |
| gaaggatttt  | gacaccctga  | aggtctatga  | cacacagctg  | gagaatgtgg  | aggcctttga  | 4920 |
| gggcctgtct  | gacttttgta  | acaccttcaa  | gctgtaccgg  | ggcaagacgc  | aggagggagc  | 4980 |
| agaagatcca  | tctgtgattg  | gtgaatttaa  | gggcctcttc  | aaaatttatc  | ccctcccaga  | 5040 |
| agaccagcc   | atccccatgc  | ccccaaagaca | gttccaccag  | ctggccgccc  | agggacccca  | 5100 |
| ggagtgcctg  | gtccgtatct  | acattgtccg  | agcatttggc  | ctgcagccca  | aggaccccaa  | 5160 |
| tggaaagtgt  | gatcccttaca | tcaagatctc  | catagggaag  | aaatcagtg   | gtgaccagga  | 5220 |
| taactacatc  | ccctgcacgc  | tggagcccg   | atgttcgagc  | tatgacctcc  | tgacctgcac  | 5280 |
| tctgcctctg  | gagaaggacc  | taaagatcac  | tctctatgac  | ctgctgtcca  | tctccaagga  | 5340 |
| cgaaaagatc  | ggtgagacgg  | tcgtcgacct  | ggagaacagg  | aaccagtggc  | agtttggggc  | 5400 |
| tcgctgtgga  | ctcccacaga  | cctactgtgt  | ctctggaccg  |             | gggaccagct  | 5460 |

25/68

|             |             |             |             |            |             |      |
|-------------|-------------|-------------|-------------|------------|-------------|------|
| ccgccccctcc | cagctcctcc  | acctctctctg | ccagcagcat  | agagtcaagg | cacctgtgtga | 5520 |
| ccggacagac  | cgtgtaatgt  | ttcaggataa  | agaatattcc  | attgaagaga | tagaggctgg  | 5580 |
| caggatccca  | aaccacacac  | tgggcccagt  | ggaggagcgt  | ctggctctgc | atgtgcttca  | 5640 |
| gcagcagggc  | ctgggtcccgg | agcacgtgga  | gtcacggccc  | ctctacagcc | ccctgcagcc  | 5700 |
| agacatcgag  | caggggaagc  | tgcagatgtg  | ggtcgacctg  | tttccgaagg | ccctggggcg  | 5760 |
| gcctggacct  | cccttcaaca  | tcaccccacg  | gagagccaga  | aggtttttcc | tgcgttgtat  | 5820 |
| tatctggaat  | accagagatg  | tgatcctgga  | tgacctgagc  | ctcacggggg | agaagatgag  | 5880 |
| cgacatttat  | gtgaaagggt  | ggatgattgg  | ctttgaagaa  | cacaagcaaa | agacagacgt  | 5940 |
| gcattatcgt  | tccctgggag  | gtgaaggcaa  | cttcaactgg  | aggttcattt | tccccttcga  | 6000 |
| ctacctgcca  | gctgagcaag  | tctgtacctg  | tgccaagaag  | gatgccttct | ggaggctgga  | 6060 |
| caagactgag  | agcaaaatcc  | cagcacgagt  | ggtgttccag  | atctgggaca | atgacaagtt  | 6120 |
| ctcctttgat  | gattttctgg  | gctccctgca  | gctcgatctc  | aaccgcagtc | ccaagccagc  | 6180 |
| caagacagcc  | aagaagtgtc  | ccttggaaca  | gctggatgat  | gctttccacc | cagaatgggt  | 6240 |
| tgtgtccctt  | tttgagcaga  | aaacagtga   | gggctgggtg  | ccctgtgtag | cagaagaggg  | 6300 |
| tgagaagaaa  | atactggcgg  | gcaagctgga  | aatgaccttg  | gagattgtag | cagagagtgga | 6360 |
| gcatgaggag  | cggcctgctg  | gccagggccg  | agatgagccc  | aacatgaacc | ctaagcttga  | 6420 |
| ggacccaagg  | cgccccgaca  | cctccttctc  | gtgggtttacc | tccccatata | agaccatgaa  | 6480 |
| gttcatcctg  | tggcggcggt  | tccggtgggc  | catcatcctc  | ttcatcatcc | tcttcatcct  | 6540 |
| gctgctgttc  | ctggccatct  | tcatctacgc  | cttcccgaac  | tatgctgcca | tgaagctggg  | 6600 |
| gaagcccttc  | agctgaggac  | tctcctgccc  | tgtagaagg   | gccgtggggg | cccctccagc  | 6660 |
| atgggactgg  | cctgcctcct  | ccgcccagct  | cgcgagctc   | ctccagacct | cctaggcctg  | 6720 |
| attgtcctgc  | caggggtggc  | agacagacag  | atggaccggc  | ccacactccc | agagttgcta  | 6780 |
| acatggagct  | ctgagatcac  | cccacttcca  | tcatttctct  | ctcccccaac | ccaacgcttt  | 6840 |
| tttgatcag   | ctcagacata  | tttcagtata  | aaacagttgg  | aaccacaaaa | aaaaaaaaaa  | 6900 |
| aaaaaaaaaa  | a           |             |             |            |             | 6911 |

<210> 18  
 <211> 6911  
 <212> DNA  
 <213> Homo sapiens

|            |             |             |            |             |            |      |
|------------|-------------|-------------|------------|-------------|------------|------|
| <400> 18   |             |             |            |             |            |      |
| tgcaccgccc | agccagggtgc | aaaatgccgt  | gtcattggga | gactccgcag  | ccggagcatt | 60   |
| agattacagc | tgcacggagc  | tccgggaagg  | cggcgggggt | ggaagatgag  | cagaagcccc | 120  |
| tggtctcggg | acgccggctg  | acaagcgggg  | tgagcgcagg | cggggcgggg  | acccagccta | 180  |
| gcccactgga | gcagccgggg  | gtggcccgtt  | cccccttaag | agcaactgct  | ctaagccagg | 240  |
| agccagagat | tcgagccggc  | ctcgcccagc  | cagccctctc | cagcgagggg  | accacaagc  | 300  |
| ggcgccctcg | ccctcccagc  | ctttccgagc  | cctcttttgc | ccctgggcgc  | acggggccct | 360  |
| acacgcgcca | agcatgctga  | gggtcttcat  | cctctatgcc | gagaacgtcc  | acacaccgga | 420  |
| caccgacatc | agcgatgcct  | actgctccgc  | ggtgtttgca | gggggtgaaga | agagaaccaa | 480  |
| agtcatcaag | aacagcgtga  | accctgtatg  | gaatgaggga | tttgaatggg  | acctcaaggg | 540  |
| catccccctg | gaccagggtc  | ctgagcttca  | tgtggtgggc | aaagaccatg  | agacgatggg | 600  |
| gaggaacagg | ttcctggggg  | aagccaaggt  | cccactccga | gaggtcctcg  | ccacccttag | 660  |
| tctgtccgcc | agcttcaatg  | ccccctgct   | ggacaccaag | aagcagccca  | caggggcctc | 720  |
| gctggtcctg | caggtgtcct  | acacaccgct  | gctggagctg | gtgccctgtg  | tcccgcctcc | 780  |
| tactcctctg | gagccctccc  | cgactctgcc  | tgacctggat | gtagtggcag  | acacaggagg | 840  |
| agaggaagac | acagaggacc  | agggactcac  | tggagatgag | gcggagccat  | tcctggatca | 900  |
| aagcggaggc | ccggggggtc  | ccaccacccc  | aaggaaacta | ccttcacgtc  | ctccgcccc  | 960  |
| ctaccccggg | atcaaaaaga  | agcgaagtgc  | gcctacatct | agaaagctgc  | tgctcagaca | 1020 |
| accgcaggat | ttccagatca  | gggtccaggt  | gatcgagggg | cgccagctgc  | cgggggtgaa | 1080 |
| catcaagcct | gtgggtcaagg | ttaccgctgc  | agggcagacc | aagcggacgc  | ggatccacaa | 1140 |
| gggaaacagc | ccactcttca  | atgagactct  | tttcttcaac | ttgtttgact  | ctcctgggga | 1200 |
| gctgtttgat | gagcccatct  | ttatcacggt  | ggtagactct | cgcttctctc  | ggacagatgc | 1260 |
| tctcctcggg | gagttccgga  | tggacgtggg  | caccatttac | agagagcccc  | ggcacgccta | 1320 |
| tctcaggaag | tggtctgctg  | tctcagaccc  | tgatgacttc | tctgctgggg  | ccagaggcta | 1380 |
| cctgaaaaca | agcctttgtg  | tgtcggggcc  | tggggacgaa | gcgcctctgg  | agagaaaaga | 1440 |
| cccctctgaa | gacaaggagg  | acattgaaag  | caacctgtct | cggcccacag  | gcgtagccct | 1500 |
| gcgaggagcc | cacttctgcc  | tgaaggctct  | ccgggcccag | gacttgccgc  | agatggacga | 1560 |
| tgccgtgatg | gacaacgtga  | aacagatctt  | tggtctcgag | agtaacaaga  | agaacttggt | 1620 |
| ggaccctctt | gtggaggtca  | gctttgctgg  | gaaaatgctg | tgagcaaga   | tcttgagaaa | 1680 |
| gacggccaac | cctcagtggg  | accagaacat  | cacactgcct | gccatgtttc  | cctccatgtg | 1740 |
| cgaaaaaatg | aggattcgta  | tcatagactg  | ggaccgcctg | actcacaatg  | acatcgtggc | 1800 |
| taccacctac | ctgagtatgt  | cgaaaaatct  | tgcccttgga | ggagaaatag  | aagaggagcc | 1860 |
| tgcaggtgct | gtcaagcctt  | cgaagccctc  | agacttggtg | gactacctgg  | gcttccctcc | 1920 |
| cacttttggg | ccctgctaca  | tcaacctcta  | tggcagtcct | agagagttca  | caggcttccc | 1980 |
| agacccttac | acagagctca  | acacaggcaa  | gggggaaggt | gtggcttatc  | gtggccggct | 2040 |
| tctgctctcc | ctggagacca  | agctgggtgga | gcacagtga  | cagaaggtgg  | aggaccttcc | 2100 |

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| tgcggatgac  | atcctccggg  | tggagaagta  | ccttaggagg  | cgcaagtact  | ccctgtttgc  | 2160 |
| ggccttctac  | tcagccacca  | tgtgagga    | tgtggatgat  | gccatccagt  | ttgaggtcag  | 2220 |
| catcgggaac  | tacgggaaca  | agttcgacat  | gacctgcctg  | ccgctggcct  | ccaccactca  | 2280 |
| gtacagccgt  | gcagtctttg  | acgggtgcca  | ctactactac  | ctaccctggg  | gtaacgtgaa  | 2340 |
| acctgtggtg  | gtgctgtcat  | cctactggga  | ggacatcagc  | catagaatcg  | agactcagaa  | 2400 |
| ccagctgctt  | gggattgctg  | accggctgga  | agctggcctg  | gagcagggtcc | acctggccct  | 2460 |
| gaaggcgag   | gtgtccacgg  | aggacgtgga  | ctcgctggtg  | gctcagctga  | cggatgagct  | 2520 |
| catcgcaggc  | tgcagccagc  | ctctgggtga  | catccatgag  | acaccctctg  | ccaccacact  | 2580 |
| ggaccagtac  | ctgtaccagc  | tgcgcaccca  | tcacctgagc  | caaatactg   | aggctgccct  | 2640 |
| ggccctgaag  | ctcggccaca  | gtgagctccc  | tgcagctctg  | gagcaggcg   | aggactggct  | 2700 |
| cctgctgctg  | cgtgccctgg  | cagaggagcc  | ccagaacagc  | ctgccggaca  | tcgtcatctg  | 2760 |
| gatgtgagc   | ggagacaagc  | gtgtggcata  | ccagcgggtg  | cccgccacc   | aagtcctctt  | 2820 |
| ctcccgccg   | ggtgccaaact | actgtggcaa  | gaattgtggg  | aagctacaga  | caatctttct  | 2880 |
| gaaatatccg  | atggagaagg  | tgcctggcgc  | ccggatgcca  | gtgcagatac  | gggtcaagct  | 2940 |
| gtggtttggg  | ctctctgtgg  | atgagaagga  | tttcaaccag  | tttgcctgag  | ggaagctgtc  | 3000 |
| tgtctttgct  | gaaacctatg  | agaacgagac  | taagtggcc   | cttgttggga  | actggggcac  | 3060 |
| aacgggctc   | acctaccca   | agttttctga  | cgtcacgggc  | aagatcaagc  | tacccaagga  | 3120 |
| cagcttccgc  | ccctcgcccg  | gctggacctg  | ggctggagat  | tggttcgtgt  | gtccggagaa  | 3180 |
| gactctgctc  | catgacatgg  | acgccgtca   | cctgagcttc  | gtggaagagg  | tggttgagaa  | 3240 |
| ccagaccgg   | cttcccgag   | gccagtggat  | ctacatgagt  | gacaactaca  | ccgatgtgaa  | 3300 |
| cggggagaag  | gtgcttccca  | aggatgacat  | tgagtggcca  | ctgggctgga  | agtgggaaga  | 3360 |
| tgaggaatgg  | tccacagacc  | tcaaccgggc  | tgctgatgag  | caaggctggg  | agtatagcat  | 3420 |
| caccatcccc  | ccggagcggg  | agccgaagca  | ctgggtccct  | gctgagaaga  | tgtactacac  | 3480 |
| acaccgacgg  | cggcgctggg  | tgcgcctgcg  | caggagggat  | ctcagccaaa  | tggaaagcact | 3540 |
| gaaaaggcac  | aggcaggcgg  | aggcggaggg  | cgagggtgg   | gagtacgcct  | ctctttttgg  | 3600 |
| ctggaagtct  | accctcgagt  | accgcaagac  | agatgccttc  | cgccgcggcc  | gctggcgccg  | 3660 |
| tcgcatggag  | ccactggaga  | agacggggcc  | tgcagctgtg  | tttgcccttg  | agggggccct  | 3720 |
| gggcggcgtg  | atggatgaca  | agagtgaaga  | ttccatgtcc  | gtctccacct  | tgagcttcgg  | 3780 |
| tgtgaacaga  | cccacgattt  | cctgcatatt  | cgactatggg  | aaccgctacc  | atctacgctg  | 3840 |
| ctacatgtac  | caggcccggg  | acctggctgc  | gatggacaag  | gactcttttt  | ctgactcccta | 3900 |
| tgccatgctc  | tcttctctgc  | accagagcca  | gaagacgggtg | gtggtgaaga  | acacccttaa  | 3960 |
| ccccacctgg  | gaaccagacg  | tcattcttcta | cgagatcgag  | atctttggcg  | agccggccac  | 4020 |
| agttgctgag  | caaccgcccc  | gcattgtggt  | ggagctgtac  | gaccatgaca  | cttatggtgc  | 4080 |
| agacgagttt  | atgggtcgct  | gcattctgtca | accgagctcg  | gaacggatgc  | cacggctggc  | 4140 |
| ctggttccca  | gtgaacgggg  | gcagccagcc  | ctcgggggag  | ctgctggcct  | ctttttagct  | 4200 |
| catccagaga  | cagaagccgg  | ccatccacca  | tattcctggt  | tttgaggtgc  | aggagacatc  | 4260 |
| aagggtcctg  | gatgagctcg  | aggacacaga  | cctgccctac  | ccaccacccc  | agagggaggc  | 4320 |
| caacatctac  | atgggtccctc | agaacatcaa  | gccagcgctc  | cagcgtaccg  | ccatcgagat  | 4380 |
| cctggcatgg  | ggcctgcgga  | acatgaagag  | ttaccagctg  | gccaacatct  | cctccccag   | 4440 |
| cctcgtggta  | gagtgtgggg  | gccagacggg  | gcagtcctgt  | gtcatcagga  | acctccggaa  | 4500 |
| gaaccccaac  | tttgacatct  | gcaccctctt  | catggaagtg  | atgctgcccc  | gggaggagct  | 4560 |
| ctactgcccc  | cccatcaccg  | tcaaggtcat  | cgataaccgc  | cagtttgggc  | gccggcctgt  | 4620 |
| ggtgggcccag | tgtaccatcc  | gctccctgga  | gagcttccctg | tgtgaccctt  | actcggcgga  | 4680 |
| gagtccatcc  | ccacagggtg  | gcccagacga  | tgtgagcccta | ctcagtcctg  | gggaagacgt  | 4740 |
| gctcatcgac  | atgatgaca   | cattcccccct | ctatccccatc | caggaggaag  | agttcatcga  | 4800 |
| ttggtggagc  | aaattctttg  | cctccatagg  | ggagagggaa  | aagtgcggct  | cctacctgga  | 4860 |
| gaaggatttt  | gacaccctga  | aggtctatga  | cacacagctg  | gagaatgtgg  | aggcctttga  | 4920 |
| gggcctgtct  | gacttttgta  | acaccttcaa  | gctgtaccgg  | ggcaagacgc  | aggaggagac  | 4980 |
| agaagatcca  | tctgtgattg  | gtgaatttaa  | gggcctcttc  | aaaattttatc | ccctcccaga  | 5040 |
| agaccagcc   | atccccatgc  | ccccaaagaca | gttccaccag  | ctggccgccc  | agggacccca  | 5100 |
| ggagtgcctg  | gtccgtatct  | acattgtccg  | agcatttggc  | ctgcagccca  | aggaccccaa  | 5160 |
| tggaaagtgt  | gatccttaca  | tcaagatctc  | catagggaag  | aaatcagtga  | gtgaccagga  | 5220 |
| taactacatc  | ccctgcacgc  | tggagcccgt  | atttggaag   | atgttcgagc  | tgacctgcac  | 5280 |
| tctgcccctg  | gagaaggacc  | tctctatgac  | tctctatgac  | tatgacctcc  | tctccaagga  | 5340 |
| cgaaaagatc  | ggtgagacgg  | tcgtcgacct  | ggagaacagg  | ctgctgtcca  | agtttggggc  | 5400 |
| tcgctgtgga  | ctcccacaga  | cctactgtgt  | ctctggaccg  | aaccagtggc  | gggaccagct  | 5460 |
| ccgcccctcc  | cagtcctcc   | acctcttctg  | ccagcagcat  | agagtcaagg  | cacctgtgta  | 5520 |
| ccggacagac  | cgtgtaatgt  | ttcaggataa  | agaatattcc  | attgaagaga  | tagaggctgg  | 5580 |
| caggatccca  | aacccacacc  | tgggcccagt  | ggaggagcgt  | ctggctctgc  | atgtgcttca  | 5640 |
| gcagcagggc  | ctgttcccg   | agcacgtgga  | gtcacggccc  | ctctacagcc  | ccctgcagcc  | 5700 |
| agacatcgag  | caggggaagc  | tgcagatgtg  | ggtcgacctta | tttccgaagg  | ccctggggcg  | 5760 |
| gcctggacct  | cccttcaaca  | tcacccacag  | gagagccaga  | aggtttttcc  | tgcttgtat   | 5820 |
| tatctggaat  | accagagatg  | tgatcctgga  | tgacctgagc  | ctcacggggg  | agaagatgag  | 5880 |
| gcacatttat  | gtgaaagggt  | ggatgattgg  | ctttgaagaa  | cacaagcaaa  | agacagacgt  | 5940 |
| gcattatcgt  | tccctgggag  | gtgaaggcaa  | cttcaactgg  | aggttcattt  | tccccttcga  | 6000 |
| ctacctgcca  | gctgagcaag  | tctgtaccat  | tgccaagaag  | gatgccttct  | ggaggctgga  | 6060 |
| caagactgag  | agcaaaatcc  | cagcacgagt  | ggtgttccag  | atctgggaca  | atgacaagtt  | 6120 |

27/68

|            |             |            |            |            |            |      |
|------------|-------------|------------|------------|------------|------------|------|
| ctcctttgat | gattttctgg  | gctccctgca | gctcgatctc | aaccgcatgc | ccaagccagc | 6180 |
| caagacagcc | aagaagtgt   | ccttggacca | gctggatgat | gctttccacc | cagaatggtt | 6240 |
| tgtgtccctt | tttgagcaga  | aaacagtga  | gggctggtgg | ccctgtgtag | cagaagaggg | 6300 |
| tgagaagaaa | atactggcgg  | gcaagctgga | aatgaccttg | gagattgtag | cagagagtga | 6360 |
| gcatgaggag | cgccctgtctg | gccagggccg | ggatgagccc | aacatgaacc | ctaagcttga | 6420 |
| ggacccaagg | cgccccgaca  | cctccttcct | gtggtttacc | tccccataca | agaccatgaa | 6480 |
| gttcatectg | tggcggtgtt  | tccggtgggc | catcatcctc | ttcatcatcc | tcttcatect | 6540 |
| gctgctgttc | ctggccatct  | tcatctacgc | cttcccgaa  | tatgctgcca | tgaagctggt | 6600 |
| gaagcccttc | agctgaggac  | tctcctgccc | tgtagaaggg | gccgtggggg | cccctccagc | 6660 |
| atgggactgg | cctgcctcct  | ccgcccagct | cggcgagctc | ctccagacct | cctaggcctg | 6720 |
| attgtcctgc | cagggtgggc  | agacagacag | atggaccggc | ccacactccc | agagttgcta | 6780 |
| acatggagct | ctgagatcac  | cccacttcca | tcatttcctt | ctcccccaac | ccaacgcttt | 6840 |
| tttgatcag  | ctcagacata  | tttcagtata | aaacagttgg | aaccacaaaa | aaaaaaaaaa | 6900 |
| aaaaaaaaaa | a           |            |            |            |            | 6911 |

<210> 19  
 <211> 6911  
 <212> DNA  
 <213> Homo sapiens

|             |             |             |            |             |             |      |
|-------------|-------------|-------------|------------|-------------|-------------|------|
| tcgaccgccc  | agccaggtgc  | aaaatgccgt  | gtcattggga | gactccgcag  | ccggagcatt  | 60   |
| agattacagc  | tcgacggagc  | tcgggaaggg  | cgccgggggt | ggaagatgag  | cagaagcccc  | 120  |
| tgttctcgga  | acgccggctg  | acaagcgggg  | tgagcgcagg | cggggcgggg  | acccagccta  | 180  |
| gcccactgga  | gcagccgggg  | gtggcccggt  | cccccttaag | agcaactgct  | ctaagcctag  | 240  |
| agccagagat  | tcgagccggc  | ctcgcccagc  | cagccctctc | cagcgagggg  | acccacaagc  | 300  |
| ggcgccctcg  | ccctcccagc  | ctttccgagc  | cctctttgcg | ccctggggcg  | acggggccct  | 360  |
| acacgcgcca  | agcatgctga  | gggtcttcat  | cctctatgcc | gagaacgtcc  | acacacccga  | 420  |
| caccgacatc  | agcgtgcctt  | actgctccgc  | gggtgttgca | ggggtgaaga  | agagaaccaa  | 480  |
| agtcatacaag | aacagcgtga  | accctgtatg  | gaatgagggg | tttgaatggg  | acctcaaggg  | 540  |
| catccccctg  | gaccagggct  | ctgagcttca  | tgtgggtggt | aaagaccatg  | agacgatggg  | 600  |
| gaggaacagg  | ttcctggggg  | aagccaaggt  | cccactccga | gaggtcctcg  | ccacccttag  | 660  |
| tctgtccgcc  | agcttcaatg  | ccccctgct   | ggacaccaag | aagcagccca  | caggggcctc  | 720  |
| gctggtcctg  | caggtgtcct  | acacaccgct  | gcctggagct | gtgcccctgt  | tcccggcccc  | 780  |
| tactcctctg  | gagccctccc  | cgactctgcc  | tgacctggat | gtagtggcag  | acacaggagg  | 840  |
| agaggaagac  | acagaggacc  | agggactcac  | tggagatgag | gcggagccat  | tcctggatca  | 900  |
| aagcggaggc  | ccggggggctc | ccaccacccc  | aaggaaacta | ccttcacgtc  | ctccgccccca | 960  |
| ctaccccggg  | atcaaaagaa  | agcgaagtgc  | gcctacatct | agaaagctgc  | tgtcagacaa  | 1020 |
| accgcaggat  | ttccagatca  | gggtccagggt | gatcgagggg | cgccagctgc  | cggggggtgaa | 1080 |
| catcaagcct  | gtggtcaagg  | ttaccgtctc  | agggcagacc | aagcggacgc  | ggatccacaa  | 1140 |
| gggaaacagc  | ccactcttca  | atgagactct  | tttcttcaac | ttgtttgact  | ctcctgggga  | 1200 |
| gctgtttgat  | gagcccatct  | ttatcacggg  | ggtagactct | cgttctctca  | ggacagatgc  | 1260 |
| tctcctcggg  | gagttccgga  | tggacgtggg  | caccatttca | agagagcccc  | ggcacgccta  | 1320 |
| tctcaggaag  | tggctgctgc  | tctcagaccc  | tgatgacttc | tctgctgggg  | ccagaggcta  | 1380 |
| cctgaaaaca  | agcctttgtg  | tgctggggcc  | tggggacgaa | gcgcctctgg  | agagaaaaga  | 1440 |
| cccctctgaa  | gacaaggagg  | acattgaaag  | caacctgctc | cgcccccacag | gcgtagccct  | 1500 |
| gcgaggagcc  | cacttctgcc  | tgaaggtctt  | ccgggcccag | gacttgccgc  | agatggacga  | 1560 |
| tgcctgtgat  | gacaacgtga  | aacagatctt  | tggcttcgag | agtaacaaga  | agaacttggt  | 1620 |
| ggaccccttt  | gtggaggtca  | gctttgcggg  | gaaaatgctg | tgcagcaaga  | tcttggagaa  | 1680 |
| cagggccaac  | cctcagtggg  | accagaacat  | cacactgcct | gccatgtttc  | cctccatgtg  | 1740 |
| cgaaaaaatg  | aggattcgta  | tcatagactg  | ggacgcctg  | actcacaatg  | acatcgtggc  | 1800 |
| taccacctac  | ctgagtatgt  | cgaaaatctc  | tgcccctgga | ggagaaatag  | aagaggagcc  | 1860 |
| tgcagggtgct | gtcaagcctt  | cgaaaagcctc | agacttgatg | gactacctgg  | gcttcctccc  | 1920 |
| cactttttggg | ccctgctaca  | tcaacctcta  | tggcagtcct | agagagttca  | caggcttccc  | 1980 |
| agacccttac  | acagagctca  | acacaggcaa  | gggggaaggt | gtggcttatc  | gtggccggct  | 2040 |
| tctgctctcc  | ctggagacca  | agctggtgga  | gcacagtga  | cagaaggtgg  | aggaccttcc  | 2100 |
| tgcggatgac  | atcctccggg  | tggagaagta  | ccttaggagg | cgcaagtact  | ccctgtttgc  | 2160 |
| ggccttctac  | tcagccacca  | tgtgacagga  | tgtggatgat | gccatccagt  | ttgaggtcag  | 2220 |
| catcggaac   | tacgggaaca  | agttcgacat  | gacctgcctg | ccgctggcct  | ccaccactca  | 2280 |
| gtacagccgt  | gcagtctttg  | acgggtgccca | ctactactac | ctaccctggg  | gtaacgtgaa  | 2340 |
| acctgtgggtg | gtgctgtcat  | cctactggga  | ggacatcagc | catagaatcg  | agactcagaa  | 2400 |
| ccagctgctt  | gggattgctg  | accggctgga  | agctggcctg | gagcaggtcc  | acctggccct  | 2460 |
| gaaggcgcag  | tgtctccacg  | aggacgtgga  | ctcgctgggt | gctcagctga  | cggatgagct  | 2520 |
| catcgagagc  | tgcagccagc  | ctctgggtga  | cctccatgag | acacctctg   | ccaccacct   | 2580 |
| ggaccagtac  | ctgtaccagc  | tgcgcaccca  | tcacctgagc | caaactactg  | aggctgcctt  | 2640 |
| ggccctgaag  | ctcggccaca  | gtgagctccc  | tgcagctctg | gagcagggcg  | aggactggct  | 2700 |
| cctgctctg   | cgtgccctgg  | cagaggagcc  | ccagaacagc | ctgccggaca  | tcgtcatctg  | 2760 |

|             |             |             |             |            |             |      |
|-------------|-------------|-------------|-------------|------------|-------------|------|
| gatgctgcag  | ggagacaagc  | gtgtggcata  | ccagcgggtg  | cccgccacc  | aagtcctctt  | 2820 |
| ctcccggcgg  | ggtgccaaact | actgtggcaa  | gaattgtggg  | aagctacaga | caatctttct  | 2880 |
| gaaatatccg  | atggagaagg  | tgcctggcgc  | ccggatgcca  | gtgcagatac | gggtcaagct  | 2940 |
| gtggtttggg  | ctctctgtgg  | atgagaagga  | gttcaaccag  | tttgctgagg | ggaagctgtc  | 3000 |
| tgtctttgct  | gaaacctatg  | agaacgagac  | taagttggcc  | cttggtggga | actggggcac  | 3060 |
| aacgggcctc  | acctaccca   | agttttctga  | cgtcacgggc  | aagatcaagc | taccaagga   | 3120 |
| cagcttcgcg  | ccctcggccg  | gctggacctg  | ggctggagat  | tggttcgtgt | gtccggagaa  | 3180 |
| gactctgctc  | catgacatgg  | acgccgggtc  | cctgagcttc  | gtggaagagg | tgtttgagaa  | 3240 |
| ccagacccgg  | cttcccggag  | gccagtggat  | ctacatgagt  | gacaactaca | ccgatgtgaa  | 3300 |
| cggggagaa   | gtgcttccca  | aggatgacat  | tgagtggcca  | ctgggctgga | agtgggaaga  | 3360 |
| tgaggaatgg  | tccacagacc  | tcaaccgggc  | tgtcgatgag  | caaggctggg | agtatagcat  | 3420 |
| caccatcccc  | ccggagcgga  | agccgaagca  | ctgggtccct  | gctgagaaga | tgtactacac  | 3480 |
| acaccgacgg  | cggcgtctgg  | tgcgcctgcg  | caggagggat  | ctcagccaaa | tgggaagcact | 3540 |
| gaaaaggcac  | aggcaggcgg  | aggcggaggg  | cgagggctgg  | gagtacgcct | ctcttttttg  | 3600 |
| ctggaagttc  | ccactcgagt  | accgcaagac  | agatgccttc  | cgccgcgcgc | gctggcgccg  | 3660 |
| tgcgatggag  | ccactggaga  | agacggggcc  | tgcagctgtg  | tttgcccttg | agggggccct  | 3720 |
| ggcggcggtg  | atggatgaca  | agagtgaaga  | ttccatgtcc  | gtctccacct | tgagcttcgg  | 3780 |
| tgtgaacaga  | cccacgattt  | cctgcatatt  | cgactatggg  | aaccgctacc | atctacgctg  | 3840 |
| ctacatgtac  | caggcccggg  | acctggctgc  | gatggacaag  | gactcttttt | ctgatcccta  | 3900 |
| tgccatcgte  | tccttcctgc  | accagagcca  | gaagacgggt  | gtggtgaaga | acacccttaa  | 3960 |
| ccccacctgg  | gaccagacgc  | tcactcttcta | cgagatcgag  | atctttggcg | agccggccac  | 4020 |
| agttgctgag  | caaccgcccc  | gcattgtggg  | ggagctgtac  | gaccatgaca | cttatggtgc  | 4080 |
| agacgagttt  | atgggtcgct  | gcactctgtca | accgagtctg  | gaacggatgc | cacggctggc  | 4140 |
| ctggttccca  | ctgacgaggg  | gcagccagcc  | gtcgggggag  | ctgctggcct | cttttgagct  | 4200 |
| catccagaga  | gagaagccgg  | ccatccacca  | tattcctggt  | tttgaggtgc | aggagacatc  | 4260 |
| aaggatcctg  | gatgagtcctg | aggacacaga  | cctgccctac  | ccaccacccc | agaggagggc  | 4320 |
| caacatctac  | atggttcctc  | agaacatcaa  | gccagcgctc  | cagcgtaccg | ccatcgagat  | 4380 |
| cctggcatgg  | ggcctgcgga  | acatgaagag  | ttaccagctg  | gccaacatct | cctccccag   | 4440 |
| cctcgtggta  | gagtgtgggg  | gccagacggg  | gcagtctgtg  | gtcatcagga | acctccggaa  | 4500 |
| gaaccccaac  | tttgacatct  | gcaccctctt  | catggaagtg  | atgctgcccc | gggaggagct  | 4560 |
| ctactgcccc  | ctactcaccg  | tcaaggctcat | cgataaccgc  | cagtttggcc | ggcggcctgt  | 4620 |
| ggtgggcccag | tgtaccatcc  | gctccctgga  | gagcttctctg | tgtgacccct | actcggcgga  | 4680 |
| gagtccatcc  | ccacagggtg  | gcccagacga  | tgtgagccta  | ctcagtcctg | gggaagacgt  | 4740 |
| gctcatcgac  | attgatgaca  | aggagcccct  | catccccatc  | caggaggaag | agttcatcga  | 4800 |
| ttggtggagg  | aaattctttg  | cctccatagg  | ggagaggga   | aagtgcggct | cctacctgga  | 4860 |
| gaaggatttt  | gacacctga   | aggtctatga  | cacacagctg  | gagaatgtgg | aggcctttga  | 4920 |
| gggcctgtct  | gacttttgtta | acaccttcaa  | gctgtaccgg  | ggcaagacgc | aggaggagac  | 4980 |
| agaagatcca  | tctgtgattg  | gtgaatttaa  | gggcctcttc  | aaaatttatc | ccctcccaga  | 5040 |
| agaccagacc  | atccccatgc  | ccccaaagaca | gttccaccag  | ctggccgccc | agggacccca  | 5100 |
| ggagtgcctg  | gtccgtatct  | acattgtccg  | agcatttggc  | ctgcagccca | aggaccccaa  | 5160 |
| tggaaagctg  | gtcccttaca  | tcaagatctc  | catagggaag  | aaatcagtga | gtgaccagga  | 5220 |
| taactacatc  | ccctgcacgc  | tggagcccgt  | atttggaag   | atgttcgagc | tgacctgcac  | 5280 |
| tctgcctctg  | gagaaggacc  | taaagatcac  | tctctatgac  | tatgacctcc | tctccaagga  | 5340 |
| cgaaaagatc  | ggtgagacgg  | tcgctgacct  | ggagaacagg  | ctgctgtcca | agtttggggc  | 5400 |
| tcgctgtgga  | ctcccacaga  | cctactgtgt  | ctctggaccg  | aaccagtggc | gggaccagct  | 5460 |
| ccgcccctcc  | cagtcctcc   | acctctctctg | ccagcagcat  | agagtcaagg | cacctgtgta  | 5520 |
| ccggacagac  | cgtgtaatgt  | ttcaggataa  | agaatattcc  | attgaagaga | tagaggctgg  | 5580 |
| caggatccca  | aaccacaccc  | tgggcccagt  | ggaggagcgt  | ctggctctgc | atgtgcttca  | 5640 |
| gcagcagggc  | ctgggtcccgg | agcacgtgga  | gtcacggccc  | ctctacagcc | ccctgcagcc  | 5700 |
| agacatcgag  | caggggaagc  | tgcagatgtg  | ggtcgacctt  | tttccgaagg | ccctggggcg  | 5760 |
| gacctggacct | cccttcaaca  | tcaccccacg  | gagagccaga  | agggttttcc | tgcgttgtat  | 5820 |
| tatctggaat  | accagagatg  | tgatcctgga  | tgacctgagc  | ctcacggggg | agaagatgag  | 5880 |
| cgacatttat  | gtgaaagggt  | ggatgattgg  | ctttgaagaa  | cacaagcaaa | agacagacgt  | 5940 |
| gcattatcgt  | tccttgggag  | gtgaaggcaa  | cttcaactgg  | aggttcattt | tccccctcga  | 6000 |
| ctacctgcca  | gctgagcaag  | tctgtacctat | tgccaagaag  | gatgccttct | ggaggctgga  | 6060 |
| caagactgag  | agcaaaaatcc | cagcacgagt  | ggtgttccag  | atctgggaca | atgacaagtt  | 6120 |
| ctcctttgat  | gattttctgg  | gctccctgca  | gctcgatctc  | aaccgcatgc | ccaagccagc  | 6180 |
| caagacagcc  | aagaagtgt   | ccttggacca  | gctggatgat  | gctttccacc | cagaatgggt  | 6240 |
| tgtgtccctt  | tttgagcaga  | aaacagtga   | gggctggtgg  | cctgtgttag | cagaagaggg  | 6300 |
| tgagaagaaa  | atactggcgg  | gcaagctgga  | aatgaccttg  | gagattgtag | cagagagtga  | 6360 |
| gcatgaggag  | cggcctgctg  | gccagggccg  | ggatgagccc  | aacatgaacc | ctaagcttga  | 6420 |
| ggacccaagg  | cgccccgaca  | cctccttccct | gtggtttacc  | tccccataca | agaccatgaa  | 6480 |
| gttcacacctg | tggcggtgtt  | tccggtgggc  | catcatcctc  | ttcatcatcc | tcttcatcct  | 6540 |
| gctgctgttc  | ctggccatct  | tcactctacgc | cttcccgaac  | tatgctgcca | tgaagctggg  | 6600 |
| gaagcccttc  | agctgaggac  | tctcctggcc  | tgtagaaggg  | gccgtggggg | cccctccagc  | 6660 |
| atgggactgg  | cctgcctcct  | ccgcccagct  | cggcgagctc  | ctccagacct | cctaggcctg  | 6720 |
| attgtcctgc  | cagggtgggc  | agacagacag  | atggaccggc  | ccacactccc | agagttgcta  | 6780 |

29/68

|            |            |            |            |            |            |      |
|------------|------------|------------|------------|------------|------------|------|
| acatggagct | ctgagatcac | cccacttcca | tcatttctct | ctcccccaac | ccaacgcttt | 6840 |
| tttggatcag | ctcagacata | tttcagtata | aaacagttgg | aaccacaaaa | aaaaaaaaaa | 6900 |
| aaaaaaaaaa | a          |            |            |            |            | 6911 |

<210> 20  
 <211> 6911  
 <212> DNA  
 <213> Homo sapiens

<400> 20

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| tcgaccgccc  | agccaggtgc  | aaaatgccgt  | gtcattggga  | gactccgcag  | ccggagcatt  | 60   |
| agattacagc  | tcgacggagc  | tcgggaaggg  | cggggggggt  | ggaagatgag  | cagaagcccc  | 120  |
| tgttctcggg  | acgccggctg  | acaagcgggg  | tgagcgcagg  | cggggcccggg | accagccta   | 180  |
| gcccactgga  | gcagccgggg  | gtggcccgtt  | cccctttaag  | agcaactgct  | ctaagccagg  | 240  |
| agccagagat  | tcgagccggc  | ctcgcgccagc | cagccctctc  | cagcaggggg  | accacaagg   | 300  |
| ggcgccctcg  | ccctcccgac  | ctttccgagc  | cctcttttgcg | ccctgggccc  | acggggccct  | 360  |
| acacgcgcca  | agcatgctga  | gggtcttcat  | cctctatgcc  | gagaacgtcc  | acacaccgga  | 420  |
| caccgacatc  | agcgtatgct  | actgctccgc  | ggtgtttgca  | ggggtgaaga  | agagaaccaa  | 480  |
| agtcatacaag | aacagcgtga  | accctgtatg  | gaatgaggga  | tttgaatggg  | acctcaaggg  | 540  |
| catccccctg  | gaccagggct  | ctgagcttca  | tgtggtggtc  | aaagaccatg  | agacgatggg  | 600  |
| gaggaacagg  | ttcctggggg  | aagccaaggt  | cccactccga  | gaggtcctcg  | ccacccttag  | 660  |
| tctgtccgcc  | agcttcaatg  | ccccctgct   | ggacaccaag  | aagcagccca  | caggggcctc  | 720  |
| gctggtcctg  | caggtgtcct  | acacaccgct  | gcctggagct  | gtgcccctgt  | tcccgcccc   | 780  |
| tactcctctg  | gagccctccc  | cgactctgcc  | tgacctggat  | gtagtggcag  | acacaggagg  | 840  |
| agaggaagac  | acagaggacc  | agggactcac  | tggagatgag  | gcggagccat  | tcctggatca  | 900  |
| aagcggaggc  | ccggggggctc | ccaccacccc  | aaggaaacta  | ccttcacgtc  | ctccgcccc   | 960  |
| ctaccccggg  | atcaaaaaga  | agcgaagtgc  | gcctacatct  | agaaagctgc  | tgtagacaaa  | 1020 |
| accgcaggat  | ttccagatca  | gggtccagggt | gatcgagggg  | cgccagctgc  | cgggggtgaa  | 1080 |
| catcaagcct  | gtggtcaagg  | ttaccgctgc  | agggcagacc  | aagcggacgc  | ggatccacaa  | 1140 |
| gggaaacagc  | ccactcttca  | atgagactct  | tttcttcaac  | ttgtttgact  | ctcctgggga  | 1200 |
| gctgtttgat  | gagcccatct  | ttatcacgggt | ggtagactct  | cggttctctca | ggacagatgc  | 1260 |
| tctcctcggg  | gagttccgga  | tggacgtggg  | caccatttac  | agagagcccc  | ggcacgccta  | 1320 |
| tctcaggaag  | tggctgctgc  | tctcagaccc  | tgatgacttc  | tctgctgggg  | ccagaggcta  | 1380 |
| cctgaaaaca  | agcctttgtg  | tgctggggcc  | tggggacgaa  | gcgcctctgg  | agagaaaaga  | 1440 |
| cccctctgaa  | gacaaggagg  | acattgaaag  | caacctgctc  | cgggccacag  | cgctagccct  | 1500 |
| gcgaggaacc  | tacttctgcc  | tgaaggctct  | ccgggcccag  | gacttgccgc  | agatggacga  | 1560 |
| tgccgtgatg  | gacaacgtga  | aacagatctt  | tggcttcgag  | agtaacaaga  | agaacttggt  | 1620 |
| ggaccccttt  | gtggagggtca | gctttgcggg  | gaaaatgctg  | tgtagcaaga  | tcttgagaaa  | 1680 |
| gacggccaac  | cctcagtgga  | accagaacat  | cacactgcct  | gccatgtttc  | cctccatgtg  | 1740 |
| cgaaaaaatg  | aggattcgta  | tcatagactg  | ggaccgcctg  | actcacaatg  | acatcgtggc  | 1800 |
| taccacctac  | ctgagtatgt  | cgaaaatctc  | tgcccctgga  | ggagaaatag  | aagaggagcc  | 1860 |
| tgcaggtgct  | gtcaagcctt  | cgaaagcctc  | agacttggtg  | gactacctgg  | gcttccctcc  | 1920 |
| cacttttggtg | ccctgctaca  | tcaacctcta  | tggcagtcct  | agagagttca  | caggcttccc  | 1980 |
| agacccctac  | acagagctca  | acacaggcaa  | gggggaaggt  | gtggcttatc  | gtggccggct  | 2040 |
| tctgctctcc  | ctggagacca  | agctggtgga  | gcacagtga   | cagaaggtgg  | aggactttcc  | 2100 |
| tgcggatgac  | atcctccggg  | tggagaagta  | ccttaggagg  | cgcaagtact  | ccctgtttgc  | 2160 |
| ggccttctac  | tcagccacca  | tgtgcagga   | tgtggatgat  | gccatccagt  | ttgaggtcag  | 2220 |
| catcggaac   | tacgggaaca  | agttcgacat  | gacctgcctg  | ccgctggcct  | ccaccactca  | 2280 |
| gtacagccgt  | gcagtctttg  | acgggtgcc   | ctactactac  | ctaccctggg  | gtaacgtgaa  | 2340 |
| acctgtgggtg | gtgctgtcat  | cctactggga  | ggacatcagc  | catagaatcg  | agactcagaa  | 2400 |
| ccagctgctt  | gggattgctg  | accggctgga  | agctggcctg  | gagcaggtcc  | acctggccct  | 2460 |
| gaaggcgcag  | tgctccacgg  | aggacgtgga  | ctcgctgggtg | gctcagctga  | cggtatgagct | 2520 |
| catcgaggc   | tgtagccagc  | ctctgggtga  | catccatgag  | acaccctctg  | ccaccacact  | 2580 |
| ggaccagtac  | ctgtaccagc  | tgcgcaccca  | tcacctgagc  | caaactcactg | aggctgccct  | 2640 |
| ggccctgaag  | ctcgccaca   | gtgagctccc  | tgtagctctg  | gagcaggcgg  | aggactggct  | 2700 |
| cctgcgtctg  | cggtccctgg  | cagaggagcc  | ccagaacagc  | ctgcccggaca | tcgtcatctg  | 2760 |
| gatgctgcag  | ggagacaagc  | gtgtggcata  | ccagcgggtg  | cccggccacc  | aagtcctctt  | 2820 |
| ctcccgggcg  | ggtgccaaact | actgtggcaa  | gaattgtggg  | aagctacaga  | caatctttct  | 2880 |
| gaaatatccg  | atggagaagg  | tgccctggcg  | ccggatgcc   | gtgcagatac  | gggtcaagct  | 2940 |
| gtggtttggt  | ctctctgtgg  | atgagaagga  | gttcaaccag  | tttgctgagg  | ggaagctgtc  | 3000 |
| tgtctttgct  | gaaacctatg  | agaacgagac  | taagttggcc  | cttggtggga  | actggggcac  | 3060 |
| aacgggcctc  | acctacccca  | agttttctga  | cgtcacgggc  | aagatcaagc  | taccacaagga | 3120 |
| cagcttccgc  | ccctcgggcg  | gctggacctg  | ggctggagat  | tggttcgtgt  | gtccggagaa  | 3180 |
| gactctgctc  | catgacatgg  | acgcccgtca  | cctgagcttc  | gtggaagagg  | tggttgagaa  | 3240 |
| ccagaccctc  | cttcccggag  | gccagtggat  | ctacatgagt  | gacaactaca  | ccgatgtgaa  | 3300 |
| cgggggagaag | gtgcttccca  | aggatgacat  | tgagtgccca  | ctgggctgga  | agtgaggaga  | 3360 |
| tgaggaatgg  | tccacagacc  | tcaaccgggc  | tgtcgatgag  | caaggctggg  | agtatagcat  | 3420 |



30/68

|             |             |             |             |            |            |      |
|-------------|-------------|-------------|-------------|------------|------------|------|
| caccatcccc  | ccggagcggg  | agccgaagca  | ctgggtccct  | gctgagaaga | tgtactacac | 3480 |
| acaccgacgg  | cggcgctggg  | tgcgcctgcg  | caggagggat  | ctcagccaaa | tggaagcact | 3540 |
| gaaaaggcac  | aggcaggcgg  | aggcggaggg  | cgagggtggt  | gagtacgcct | ctcttttttg | 3600 |
| ctggaagtcc  | cacctcgagt  | accgcaagac  | agatgccttc  | cgccgcgcgc | gctggcgccg | 3660 |
| tcgcatggag  | ccactggaga  | agacggggcc  | tgcagctgtg  | tttggccttg | agggggccct | 3720 |
| gggcggcggt  | atggatgaca  | agagtgaaga  | ttccatgtcc  | gtctccacct | tgagcttcgg | 3780 |
| tgtgaacaga  | cccacgattt  | cctgcatatt  | cgactatggg  | aaccgctacc | atctacgctg | 3840 |
| ctacatgtac  | caggcccggg  | acctggctgc  | gatggacaag  | gactcttttt | ctgatcccta | 3900 |
| tgccatcgtc  | tccttcctgc  | accagagcca  | gaagacgggt  | gtggtgaaga | acacccttaa | 3960 |
| ccccacctgg  | gaccagacgc  | tcattcttcta | cgagatcgag  | atctttggcg | agccggccac | 4020 |
| agttgctgag  | caaccgcca   | gcattgtggt  | ggagctgtac  | gacctgaca  | cttatgggtg | 4080 |
| agacgagttt  | atgggtcgct  | gcatctgtca  | accgagctct  | gaacggatgc | cacggctggc | 4140 |
| ctgggtccca  | ctgacgaggg  | gcagccagcc  | gtcgggggag  | ctgctggcct | cttttgagct | 4200 |
| catccagaga  | gagaagccgg  | ccatccacca  | tattcctggt  | tttgaggtgc | aggagacatc | 4260 |
| aaggatcctg  | gatgagtctg  | aggacacaga  | cctgcccctac | ccaccacccc | agaggagggc | 4320 |
| caacatctac  | atgggttcctc | agaacatcaa  | gccagcgctc  | cagcgtaccg | ccatcgagat | 4380 |
| cctggcatgg  | ggcctgcgga  | acatgaagag  | ttaccagctg  | gccaacatct | cctccccag  | 4440 |
| cctcgtggta  | gagtgtgggg  | gccagacggg  | gcagtcctgt  | gtcatcagga | acctccggaa | 4500 |
| gaaccccaac  | tttgacatct  | gcaccctctt  | catggaagtg  | atgctgccc  | gggaggagct | 4560 |
| ctactgcccc  | cccatacccg  | tcaaggtcat  | cgataaccgc  | cagtttgccc | gccggcctgt | 4620 |
| ggtggggccag | gttaccatcc  | gctccctgga  | gagcttcctg  | tgtgacccct | actcggcgga | 4680 |
| gagtccatcc  | ccacagggtg  | gccagacga   | tgtgagccta  | ctcagtcctg | gggaagacgt | 4740 |
| gctcatcgac  | attgatgaca  | aggagcccc   | catccccatc  | caggaggaag | agttcatcga | 4800 |
| ttggtggagc  | aaattctttg  | cctccatagg  | ggagagggaa  | aagtgcggct | cctacctgga | 4860 |
| gaaggatttt  | gacaccctga  | aggtctatga  | cacacagctg  | gagaatgtgg | aggcctttga | 4920 |
| gggcctgtct  | gacttttgta  | acaccttcaa  | gctgtaccgg  | ggcaagacgc | aggaggagac | 4980 |
| agaagatcca  | tctgtgattg  | gtgaatttaa  | gggcctcttc  | aaaatttatc | ccctcccaga | 5040 |
| agaccagcc   | atccccatgc  | ccccaaagaca | gttccaccag  | ctggccgccc | agggacccca | 5100 |
| ggagtgcctg  | gtccgtatct  | acattgtccg  | agcattttgg  | ctgcagccca | aggaccccaa | 5160 |
| tggaagtgt   | gatccttaca  | tcaagatctc  | catagggaag  | aatcagtgga | gtgaccagga | 5220 |
| taactacatc  | ccctgcacgc  | tggagcccg   | atttggaaag  | atgttcgagc | tgacctgcac | 5280 |
| tctgcctctg  | gagaaggacc  | taaagatcac  | tctctatgac  | tatgacctcc | tctccaagga | 5340 |
| cgaaaagatc  | ggtgagacgg  | tcgtcgacct  | ggagaacagg  | ctgctgtcca | agtttggggc | 5400 |
| tcgctgtgga  | ctcccacaga  | cctactgtgt  | ctctggaccg  | aaccagtggc | gggaccagct | 5460 |
| ccgccccctc  | cagctccctc  | acctctctgt  | ccagcagcat  | agagtcaagg | cacctgtgta | 5520 |
| ccggacagac  | cgtgtaattg  | ttcaggataa  | agaatattcc  | attgaagaga | tagaggctgg | 5580 |
| caggatccca  | aaccacaccc  | tgggcccagt  | ggaggagcgt  | ctggctctgc | atgtgcttca | 5640 |
| gcagcagggc  | ctgggtcccg  | agcacgtgga  | gtcacggccc  | ctctacagcc | ccctgcagcc | 5700 |
| agacatcgag  | cagggggaagc | tgcagatgtg  | ggtcgacctc  | tttccgaagg | ccctggggcg | 5760 |
| gcctggacct  | cccttcaaca  | tcaccccacg  | gagagccaga  | aggtttttcc | tgcgttgtat | 5820 |
| tatctggaat  | accagagatg  | tgatcctgga  | tgacctgagc  | ctcacggggg | agaagatgag | 5880 |
| cgacatttat  | gtgaaagggt  | ggatgattgg  | ctttgaagaa  | cacaagcaaa | agacagacgt | 5940 |
| gcgttatcgt  | tccttgggag  | gtgaaggcaa  | cttcaactgg  | aggttcattt | tccccttcga | 6000 |
| ctacctgcca  | gctgagcaag  | tctgtacctc  | tgccaagaag  | gatgccttct | ggaggctgga | 6060 |
| caagactgag  | agcaaaatcc  | cagcacgagt  | ggtgttccag  | atctgggaca | atgacaagtt | 6120 |
| ctcctttgat  | gattttctgg  | gctccctgca  | gctcgatctc  | aaccgcatgc | ccaagccagc | 6180 |
| caagacagcc  | agaagtgct   | ccttggaacca | gctggatgat  | gctttccacc | cagaatgggt | 6240 |
| tgtgtccctt  | tttgagcaga  | aaacagtga   | gggctggtgg  | ccctgtgtag | cagaagaggg | 6300 |
| tgagaagaaa  | atactggcgg  | gcaagctgga  | aatgaccttg  | gagattgtag | cagagagtga | 6360 |
| gcatgaggag  | cgccctgctg  | gccagggcgg  | ggatgagccc  | aacatgaacc | ctaagcttga | 6420 |
| ggacccaagg  | cgccccgaca  | cctccttcct  | gtggtttacc  | tccccataca | agaccatgaa | 6480 |
| gttcatcctg  | tggcggcgtt  | tccggtgggc  | catcatcctc  | ttcatcatcc | tcttcatcct | 6540 |
| gctgctgttc  | ctggccatct  | tcattctacgc | cttcccgaac  | tatgctgcca | tgaagctggt | 6600 |
| gaagcccttc  | agctgaggac  | tctcctgccc  | tgtagaaggg  | gccgtggggg | cccctccagc | 6660 |
| atgggactgg  | cctgcctcct  | ccgcccagct  | cggcgagctc  | ctccagacct | cctaggcctg | 6720 |
| attgtcctgc  | cagggtgggc  | agacagacag  | atggaccggc  | ccacactccc | agagttgcta | 6780 |
| acatggagct  | ctgagatcac  | cccacttcca  | tcatttcctt  | ctcccccaac | ccaacgcttt | 6840 |
| tttgatcag   | ctcagacata  | tttcagtata  | aaacagttgg  | aaccacaaaa | aaaaaaaaaa | 6900 |
| aaaaaaaaaa  | a           |             |             |            |            | 6911 |

<210> 21  
 <211> 6909  
 <212> DNA  
 <213> Homo sapiens

<400> 21  
 tcgaccgccc agccaggtgc aaaatgccgt gtcattggga gactccgcag ccggagcatt

60



|             |             |             |             |            |             |      |
|-------------|-------------|-------------|-------------|------------|-------------|------|
| agattacagc  | tcgacggagc  | tcgggaaggg  | cggcgggggt  | ggaagatgag | cagaagcccc  | 120  |
| tggtctcgga  | acgcccggctg | acaagcgggg  | tgagcgcagg  | cggggcgggg | accagccta   | 180  |
| gcccactgga  | gcagccgggg  | gtggcccgtt  | cccctttaag  | agcaactgct | ctaagccagg  | 240  |
| agccagagat  | tcgagccggc  | ctcggcccagc | cagccctctc  | cagcgagggg | accacaagc   | 300  |
| ggcgcctcgg  | ccctcccagc  | ctttccgagc  | cctctttgcg  | ccctgggcgc | acggggccct  | 360  |
| acacgcgcga  | agcatgctga  | gggtcttcat  | cctctatgcc  | gagaacgtcc | acacaccga   | 420  |
| caccgacatc  | agcgatgcct  | actgctccgc  | gggtgttgca  | ggggtgaaga | agagaaccaa  | 480  |
| agtcatcaag  | aacagcgtga  | accctgtatg  | gaatgagggg  | tttgaatggg | acctcaaggg  | 540  |
| catccccctg  | gaccagggct  | ctgagcttca  | tgtgggtggc  | aaagaccatg | agacgatggg  | 600  |
| gaggaacagg  | ttcctggggg  | aagccaaggt  | cccactccga  | gaggtcctcg | ccacccttag  | 660  |
| tctgtccgcg  | agcttcaatg  | ccccctgct   | ggacaccaag  | aagcagccca | caggggcctc  | 720  |
| gctggctcctg | caggtgtcct  | acacaccgct  | gcctggagct  | gtgcccctgt | ccccgcccc   | 780  |
| tactcctctg  | gagccctccc  | cgactctgcc  | tgacctggat  | gtagtggcag | acacaggagg  | 840  |
| agaggaagac  | acagaggacc  | agggactcac  | tggagatgag  | gcggagccat | tcctggatca  | 900  |
| aagcggaggc  | ccggggggctc | ccaccacccc  | aaggaaacta  | ccttcacgtc | ctccgcccc   | 960  |
| ctaccccggg  | atcaaaaagaa | agcgaagtgc  | gcctacatct  | agaaagctgc | tgctcagacaa | 1020 |
| accgcaggat  | ttccagatca  | gggtccaggt  | gatcgagggg  | cgccagctgc | cggggggtgaa | 1080 |
| catcaagcct  | gtggtcaagg  | ttaccgctgc  | agggcagacc  | aagcggacgc | ggatccacaa  | 1140 |
| gggaaacagc  | ccactcttca  | atgagactct  | tttcttcaac  | ttgtttgact | ctcctgggga  | 1200 |
| gctgtttgat  | gagcccatct  | ttatcacggt  | ggtagactct  | cggtctctca | ggacagatgc  | 1260 |
| tctcctcggg  | gagttccgga  | tggacgtggg  | caccatttac  | agagagcccc | ggcacgccta  | 1320 |
| tctcaggaag  | tggctgctgc  | tctcagaccc  | tgatgacttc  | tctgctgggg | ccagaggcta  | 1380 |
| cctgaaaaca  | agcctttgtg  | tgctggggcc  | tggggacgaa  | gcgcctctgg | agagaaaaga  | 1440 |
| cccctctgaa  | gacaaggagg  | acattgaaag  | caacctgctc  | cggcccacag | gcgtagccct  | 1500 |
| gcgaggagcc  | cacttctgcc  | tgaaggctct  | ccgggcccag  | gacttgccgc | agattgacga  | 1560 |
| tgccgtgatg  | aacagatctt  | tggcttcgag  | tggcttcgag  | agtaacaaga | agaacttggt  | 1620 |
| ggaccocctt  | gtggaggtca  | gctttgcggg  | gaaaatgctg  | tgacgcaaga | tcttgagaa   | 1680 |
| gacggccaac  | cctcagtgga  | accagaacat  | cacactgcct  | gccatgtttc | cctccatgtg  | 1740 |
| cgaaaaaatg  | aggattcgtg  | tcatagactg  | ggaccgcctg  | actcacaatg | acatcgtggc  | 1800 |
| taccacctac  | ctgagtatgt  | cgaaaaatctc | agacttggtg  | ggagaaatag | aagagcagcc  | 1860 |
| tgccaggtgc  | gtcaagcctt  | cgaaagcctc  | agacttggtg  | gactacctgg | gcttcctccc  | 1920 |
| cactttttggg | ccctgctaca  | tcaacctcta  | tggcagtcct  | agagagttca | caggcttccc  | 1980 |
| agacccctac  | acagagctca  | acacaggcaa  | gggggaaggt  | gtggcttatc | gtggccggct  | 2040 |
| tctgctctcc  | ctggagacca  | agctggtgga  | gcacagtga   | cagaaggtgg | aggaccttcc  | 2100 |
| tgccggatgac | atcctccggg  | tggagaagta  | ccttaggagg  | cgcaagtact | ccctgtttgc  | 2160 |
| ggccttctac  | tcagccacca  | tgctgccagga | tgtggatgat  | gccatccagt | ttgaggtcag  | 2220 |
| catcggaac   | tacgggaaca  | agttcgacat  | gacctgcctg  | ccgctggcct | ccaccactca  | 2280 |
| gtacagccgt  | gcagtctttg  | acgggtgcca  | ctactactac  | ctaccctggg | gtaacgtgaa  | 2340 |
| acctgtggtg  | gtgctgtcat  | cctactggga  | ggacatcagc  | catagaatcg | agactcgaa   | 2400 |
| ccagctgctt  | gggattgtct  | accgctgga   | agctggcctg  | gagcaggtcc | acctggccct  | 2460 |
| gaaggcgag   | tgctccacgg  | aggcgtgga   | ctcgtggtg   | gctcagctga | cggatgagct  | 2520 |
| catcgagcgc  | tgcagccagc  | ctctgggtga  | catccatgag  | acaccctctg | ccaccacact  | 2580 |
| ggaccagtac  | ctgtaccagc  | tgcgcaccca  | tcacctgagc  | caaatactg  | aggctgccct  | 2640 |
| ggccctgaag  | ctcgccaca   | gtgagctccc  | tcgagctctg  | gagcaggcgg | aggactggct  | 2700 |
| cctgcgtctg  | cgtgcctcgg  | cagaggagcc  | ccagaacagc  | ctgccggaca | tcgtcatctg  | 2760 |
| gatgctgcag  | ggagacaagc  | gtgtggcata  | ccagcgggtg  | cccggccacc | aagtcctctt  | 2820 |
| ctcccgccgg  | ggtgccaaact | actgtggcaa  | gaattgtggg  | aagctacaga | caatctttct  | 2880 |
| gaaatatccg  | atggagaagg  | tgccctggcg  | ccggatgcca  | gtgcagatac | gggtcaagct  | 2940 |
| gtggtttggg  | ctctctgtgg  | atgagaagga  | gttcaaccag  | tttgtgagg  | ggaagctgtc  | 3000 |
| tgctcttgct  | gaaacctatg  | agaacgagac  | taagtggcc   | cctgttgagg | actggggcac  | 3060 |
| aacgggcctc  | acctacccca  | agttttctga  | cgtcacgggc  | aagatcaagc | tacccaagga  | 3120 |
| cagcttccgc  | ccctcgcccg  | gctggacctg  | ggctggagat  | tggttcgtgt | gtccggagaa  | 3180 |
| gactctgctc  | catgacatgg  | acgcccgtca  | cctgagcttc  | gtggaagagg | tgtttgagaa  | 3240 |
| ccagaccggg  | cttcccggag  | gccagtggat  | ctacatgagt  | gacaactaca | ccgatgtgaa  | 3300 |
| cggggggaag  | gtgcttccca  | aggatgacat  | tgagtggcca  | ctgggctgga | agtgggaaga  | 3360 |
| tgagggaatgg | tccacagacc  | tcaaccgggc  | tgctgatgag  | caaggctggg | agtatagcat  | 3420 |
| caccatcccc  | ccggagcggg  | agccgaagca  | ctgggtccct  | gctgagaaga | tgtactacac  | 3480 |
| acaccgacgg  | cggcgctggg  | tgcgcctgcg  | caggagggat  | ctcagccaaa | tggaaactac  | 3540 |
| gaaaaggcac  | aggcaggcgg  | aggcggaggg  | cgagggctgg  | gagtacgcct | ctctttttgg  | 3600 |
| ctggaagttc  | accctcgagt  | accgcaagac  | agatgccttc  | cgccgcggcc | gctggcgccg  | 3660 |
| tcgcatggag  | ccactggaga  | agacggggcc  | tgacgtgtg   | tttgcccttg | agggggccct  | 3720 |
| ggcgggcgtg  | atggatgaca  | agagtgaaga  | ttccatgtcc  | gtctccacct | tgagcttcgg  | 3780 |
| tgtgaacaga  | cccacgattt  | cctgcatatt  | cgactatggg  | aaccgctacc | atctacgtg   | 3840 |
| ctacatgtac  | caggcccggg  | acctggctgc  | gatggacaag  | gactcttttt | ctgatcccta  | 3900 |
| tgccatcgctc | ctcttctgc   | accagagcca  | gaagacgggtg | gtggtgaaga | acacccttaa  | 3960 |
| ccccacctgg  | gaccagacgc  | tcattcttcta | cgagatcgag  | atctttggcg | agccggccac  | 4020 |
| agttgctgag  | caaccgcccc  | gcattgtggt  | ggagctgtac  | gaccatgaca | cttatggtgc  | 4080 |

32/68

|             |             |             |            |             |             |      |
|-------------|-------------|-------------|------------|-------------|-------------|------|
| agacgagttt  | atgggtcgct  | gcattctgtca | accgagtcgt | gaacggatgc  | cacggctggc  | 4140 |
| ctggttccca  | ctgacgaggg  | gcagccagcc  | gtcgggggag | ctgctggcct  | cttttgagct  | 4200 |
| catccagaga  | gagaagccgg  | ccatccacca  | tattcctggg | tttgaggtgc  | aggagacatc  | 4260 |
| aaggatcctg  | gatgagtcgt  | aggacacaga  | cctgccctac | ccaccacccc  | agagggaggg  | 4320 |
| caacatctac  | atggttcctc  | agaacatcaa  | gccagcgctc | cagcgtaccg  | ccatcgagat  | 4380 |
| cctggcatgg  | ggcctgcgga  | acatgaagag  | ttaccagctg | gccaacatct  | cctccccag   | 4440 |
| cctcgtggta  | gagtggtggg  | gccagacggg  | gcagtcctgt | gtcatcagga  | acctccggaa  | 4500 |
| gaaccccaac  | tttgacatct  | gcaccctctt  | catggaagtg | atgctgcccc  | gggaggagct  | 4560 |
| ctactgcccc  | cccatcaccg  | tcaaggtcat  | cgataaccgc | cagtttgccc  | gccggcctgt  | 4620 |
| ggtggggccag | tgtaccatcc  | gctccctgga  | gagcttcctg | tgtgaccctt  | actcggcgga  | 4680 |
| gagtcctatcc | ccacaggggtg | gcccagacga  | tgtgagccta | ctcagtcctg  | gggaagacgt  | 4740 |
| gctcatcgac  | attgatgaca  | aggagcccc   | catccccatc | caggaggaag  | agttcatcga  | 4800 |
| ttggtggagc  | aaattccttg  | cctccatagg  | ggagagggaa | aagtgcgggt  | cctacctgga  | 4860 |
| gaaggatttt  | gacaccctga  | aggtctatga  | cacacagctg | gagaatgtgg  | aggcctttga  | 4920 |
| gggcctgtct  | gacttttgta  | acaccttcaa  | gctgtaccgg | ggcaagacgc  | aggagagac   | 4980 |
| agaagatcca  | tctgtgattg  | gtgaatttaa  | gggcctcttc | aaaattttatc | ccctcccaga  | 5040 |
| agaccagacc  | atccccatgc  | ccccaaagaca | gttccaccag | ctggccgccc  | agggacccca  | 5100 |
| ggagtgcctg  | gtccgtatct  | acattgtccg  | agcatttggt | ctgcagccca  | aggaccccaa  | 5160 |
| tggaaagtgt  | gacccctaca  | tcaagatctc  | catagggaa  | aaatcagtga  | gtgaccagga  | 5220 |
| taactacatc  | ccctgcacgc  | tggagcccg   | atttggaag  | atgttcgagc  | tgacctgcac  | 5280 |
| tctgcctctg  | gagaaggacc  | taaagatcac  | tctctatgac | tatgacctcc  | tctccaagga  | 5340 |
| cgaaaagatc  | ggtgagacgg  | tcgtcgacct  | ggagaacagg | ctgctgtcca  | agtttggggc  | 5400 |
| tcgctgtgga  | ctcccacaga  | cctactgtgt  | ctctggaccg | aaccagtggc  | gggaccagct  | 5460 |
| ccgccccctc  | cagctcctcc  | acctcttctg  | ccagcagcat | agagtcaagg  | cacctgtgta  | 5520 |
| ccggacagac  | cgtgtaatgt  | ttcaggataa  | agaatattcc | attgaagaga  | tagaggctgg  | 5580 |
| caggatccca  | aaccacaccc  | tggggccagt  | ggaggagcgt | ctggctctgc  | atgtgcttca  | 5640 |
| gcagcagggc  | ctgggtcccg  | agcacgtgga  | gtcacggccc | ctctacagcc  | ccctgcagcc  | 5700 |
| agacatcgag  | caggggaagc  | tgcagatgtg  | ggtcgacct  | tttccgaagg  | ccctggggcg  | 5760 |
| gcctggacct  | cccttcaaca  | tcacccacag  | gagagccaga | agggtttttcc | tgctgtgtat  | 5820 |
| tatctggaat  | accagagatg  | tgatcctgga  | tgacctgagc | ctcacggggg  | agaagatgag  | 5880 |
| cgacatttat  | gtgaaagggt  | ggatgattgg  | ctttgaagaa | cacaagcaaa  | agacagacgt  | 5940 |
| gcattatcgt  | tccctgggag  | gtgaaggcaa  | cttcaactgg | aggttcattt  | tccccttcga  | 6000 |
| ctacctgcca  | gctgagcaag  | tctgtaccat  | tgccaagaag | gatgccttct  | ggaggctgga  | 6060 |
| caagactgag  | caaaaatccca | gcacgagtgg  | tgttccagat | ctgggacaat  | gacaagttct  | 6120 |
| cctttgatga  | ttttctgggc  | tccctgcagc  | tcgatctcaa | ccgcatgccc  | aagccagcca  | 6180 |
| agacagccaa  | gaagtgtctc  | ttggaccagc  | tggatgatgc | tttccaccca  | gaatggtttg  | 6240 |
| tgtccctttt  | tgagcagaaa  | acagtgaagg  | gctgggtggc | ctgtgtagca  | gaagaggggtg | 6300 |
| agaagaaaat  | actggcgggc  | aagctggaaa  | tgaccttgga | gattgtagca  | gagagtgagc  | 6360 |
| atgaggagcg  | gcctgctggc  | cagggccggg  | atgagcccaa | catgaaccct  | aagcttgagg  | 6420 |
| acccaaggcg  | ccccgacacc  | tccttctctg  | ggtttacctc | ccataacaag  | accatgaagt  | 6480 |
| tcattcctgtg | gcggcggttc  | cggtgggcca  | tcattcctct | catcatcctc  | ttcatcctgc  | 6540 |
| tgctgttcct  | ggccatcttc  | atctacgcct  | tcccgaacta | tgctgccatg  | aagctggtga  | 6600 |
| agcccttcag  | ctgaggactc  | tcctgccctg  | tagaaggggc | cgtgggggtcc | cctccagcat  | 6660 |
| gggactggcc  | tgccctctcc  | gccagctcg   | gcgagctcct | ccagacctcc  | taggcctgat  | 6720 |
| tgctctgcca  | gggtgggcag  | acagacagat  | ggaccggccc | acactcccag  | agttgctaac  | 6780 |
| atggagctct  | gagatcacc   | cacttccatc  | atttctctct | cccccaacc   | aacgcttttt  | 6840 |
| tggatcagct  | cagacatatt  | tcagtataaa  | acagttggaa | ccacaaaaaa  | aaaaaaaaaa  | 6900 |
| aaaaaaaaaa  |             |             |            |             |             | 6909 |

<210> 22  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 22  
 tgggacctca agggcatccc

20

<210> 23  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens

<400> 23  
 accatgctgc aggatgtgga

20

<210> 24  
 <211> 20

33/68

<212> DNA  
 <213> Homo sapiens ;  
 <400> 24  
 gggaggtgaa ggcaacttca 20  
 <210> 25  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens  
 <400> 25  
 ctcacggggg agaagatgag 20  
 <210> 26  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens  
 <400> 26  
 ctgtggcggc gtttccggtg 20  
 <210> 27  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens  
 <400> 27  
 acatcaagga tcctggatga 20  
 <210> 28  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens  
 <400> 28  
 ctgtggcggc gtttccggtg 20  
 <210> 29  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens  
 <400> 29  
 acagacgtgc attatcgttc 20  
 <210> 30  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens  
 <400> 30  
 aagactgaga gcaaaatccc 20  
 <210> 31  
 <211> 507  
 <212> DNA  
 <213> Homo sapiens  
 <400> 31  
 tcgaccgccc agccaggtgc aaaatgccgt gtcattggga gactccgcag ccggagcatt 60  
 agattacagc tcgacggagc tcgggaaggg cggcgggggt ggaagatgag cagaagcccc 120  
 tgttctcggg acgccggctg acaagcgggg tgagcgcagg cggggcgggg acccagccta 180  
 gccactgga gcagccgggg gtggcccgtt cccctttaag agcaactgct ctaagccagg 240  
 agccagagat tcgagccggc ctgcgccagc cagccctctc cagcgagggg acccacaagc 300  
 ggcgccctcg cctcccgac ctttccgagc cctctttgcg ccctgggcgc acggggccct 360  
 acacgcgcca agcatgctga gggctctcat cctctatgcc gagaacgtcc acacaccga 420

34/68

caccgacatc agcgatgcct actgctccgc ggtgtttgca ggtaggaggg gccgaccacc 480  
ctcgccgggg tcgggggtgg gtagagg 507

<210> 32  
<211> 183  
<212> DNA  
<213> Homo sapiens

<400> 32  
aaaggcggga tgtgtctctc cattctccct tttgtgtctc ttgtaggggt gaagaagaga 60  
accaaagtca tcaagaacag cgtgaaccct gtatggaatg aggtatgtga gtttttctcc 120  
ttccttttct ctctgtctgc tgcagggggc ttgggaggag gtgccttctc agcagtgtcc 180  
ttg 183

<210> 33  
<211> 264  
<212> DNA  
<213> Homo sapiens

<400> 33  
cattcatgaa tgcctactca gtgccctggg ggcacgaagg tgaaccagac acagtctctt 60  
ctcctagagg gccataggtt aagatgcctt ttctcttttt ctccaggga tttgaatggg 120  
acctcaaggg catccccctg gaccagggtt ctgagcttca tgtgggtggc aaagaccatg 180  
agacgatggg gaggaacagg taaggtggcc agaggggggt gctccatggc ttgaaggtgc 240  
aggtaggatt gtggagtata caga 264

<210> 34  
<211> 223  
<212> DNA  
<213> Homo sapiens

<400> 34  
cagaagagcc aggggtgcctt aggctagttt tctacatttg acttctctct cctctcaggt 60  
tcctggggga agccaaggtc ccactccgag aggtcctcgc caccctagt ctgtccgcca 120  
gcttcaatgc cccctgtctg gacaccaaga agcagccac aggggtaagt gcccatcagc 180  
ctctgccagg ttaaggtcca aggcattgcc aggtggcttc ctg 223

<210> 35  
<211> 224  
<212> DNA  
<213> Homo sapiens

<400> 35  
cagtgggtccg agggcagcgc accaacctgt ccccccacgtc tcatctcttc caggcctcgc 60  
tggtcctgca ggtgtcctac acaccgtgc ctggagctgt gcccctgttc ccgcccccta 120  
ctcctctgga gccctccccg actctgcctg acctggatgt agtggcaggt gggtagccca 180  
cgttggcctg gctgggcccc agcaagaatg gccggcagtg gcac 224

<210> 36  
<211> 315  
<212> DNA  
<213> Homo sapiens

<400> 36  
aggggcaggg gcaggggccag agggccaggc ctcattaggg ccctctcctc ttagacacag 60  
gaggagagga agacacagag gaccagggac tcaactggaga tgaggcggag ccattcctgg 120  
atcaaagcgg agggccgggg gctcccacca cccaaggaa actaccttca cgtcctccgc 180  
cccactaccc cgggatcaaa agaaagcgaa gtgcgcctac atctagaaag ctgctgtcag 240  
acaaaccgca ggatttccag gtgatgaacg ggctttctct gaccccaggc tcctcttcag 300  
ccatcagctg cgggt 315

<210> 37  
<211> 249  
<212> DNA  
<213> Homo sapiens

35/68

<400> 37  
 ccagtgggtga gatgggtccct gagattttctg actctttgggg tggatgggtgg gtgggtcctta 60  
 actctttcccc cttctggctt tcagatcagg gtccagggtga tccagggggcg ccagctgccg 120  
 ggggtgaaca tcaagcctgt ggtcaagggtt accgctgcag ggcagaccaa gcggacgcgg 180  
 atccacaagg gaaacagccc actcttcaat gaggtgggag acatgggggca tgagggcaga 240  
 accttgtgg 249

<210> 38  
 <211> 185  
 <212> DNA  
 <213> Homo sapiens

<400> 38  
 ccctggcctg agggatcagc aggcactgat atgtctctct ttgctctgaa ccaacagact 60  
 cttttcttca acttggttga ctctcctggg gagctgtttg atgagcccat ctttatcacg 120  
 gtatgtctca gcagtc aaag tgttctccgt gggctgtatg tatgcacata ggtgtcagtg 180  
 cacac 185

<210> 39  
 <211> 196  
 <212> DNA  
 <213> Homo sapiens

<400> 39  
 aagagctatt ggggtggccg tgtggggccac atgtccctgt gaatgtgagc catgatcttt 60  
 ctctgcaggt ggtagactct cgttctctca ggacagatgc tctcctcggg gagttccggg 120  
 taattgctta ttttctaaaa gcagtcagtt ctcacttctc cgtgttggtg gagcctctgt 180  
 ggaccatggg cagggg 196

<210> 40  
 <211> 178  
 <212> DNA  
 <213> Homo sapiens

<400> 40  
 tggaatcgta taatgcacca cactttatctt aacgcttttg cggcaagagt ttgattttgtg 60  
 tctcctctct tgattgcaga tggacgtggg caccattttac agagagcccc gtgagttctc 120  
 accacttttg ccgtatcctt gcattttggt tctggaggct gattggggac actcattt 178

<210> 41  
 <211> 231  
 <212> DNA  
 <213> Homo sapiens

<400> 41  
 ggggtcttct gattctggga tcaccaaagg atgttgtctc tcttagggca cgcctatctc 60  
 aggaagtggc tgctgctctc agaccctgat gacttctctg ctggggccag aggctacctg 120  
 aaaacaagcc tttgtgtgct ggggcctggg gacgaagcgc ctgtgagtac atttccctgg 180  
 gtcttctcta cgtccccc cgcggcactt ggttgccggag gcaccaaacc a 231

<210> 42  
 <211> 247  
 <212> DNA  
 <213> Homo sapiens

<400> 42  
 gtcaaaaccc tgtgctcagg agcgcattgaa ggaacgtatt tggttttctt tgtagctgga 60  
 gagaaaagac ccctctgaag acaaggagga cattgaaagc aacctgctcc ggcccacagg 120  
 cgtagccctg cgaggagccc acttctgcct gaaggtcttc cgggcccagg acttgccgca 180  
 gagtgcgtgg ggcgcgccct tgggtgggag gtctgcagga ggctggaggc gcagggtctg 240  
 tgggggt 247

<210> 43  
 <211> 179  
 <212> DNA  
 <213> Homo sapiens

36/68

<400> 43  
 caggcagtga ctggtgtgtc cctcttccca gtggacgatg ccgtgatgga caacgtgaaa 60  
 cagatctttg gcttcgagag taacaagaag aacttggtgg acccctttgt ggaggtcagc 120  
 tttgcgggga aaatggtaag gagcaaggga gcaggagggt tctctcggga ggggacggg 179

<210> 44  
 <211> 202  
 <212> DNA  
 <213> Homo sapiens

<400> 44  
 ccccggggga gccagagtc cccatggagc tgatcaactt gtcccctccc tgtgtcttct 60  
 agctgtgcag caagatcttg gagaagacgg ccaaccctca gtggaaccag aacatcacac 120  
 tgctgccat ggtgagcctc ctgtccccag caaacccaag gaggcccctg gggctctggg 180  
 cttcgggagg tccagggtc ct 202

<210> 45  
 <211> 167  
 <212> DNA  
 <213> Homo sapiens

<400> 45  
 gggaggggct gttctatctt caaaaggact cttctcccaa cagcctcta ttccttctc 60  
 agtttccctc catgtgcgaa aaaatgagga ttcgtatcat agactggtga gttctgagtc 120  
 ttggagtctt tagggcgggc tgtcctgagg gggcgctccc tcagttt 167

<210> 46  
 <211> 220  
 <212> DNA  
 <213> Homo sapiens

<400> 46  
 tgtggcctga gttcctttcc tgtgtcaggc cctctctgct cccttgctct ctagggacgc 60  
 cctgactcac aatgacatcg tggctaccac ctacctgagt atgtcgaaaa tctctgcccc 120  
 tggaggagaa atagaaggta tgttccctct tcgttctgcc ctttgacccc ctgtgctctc 180  
 ccccccctcta tccagcttac acttctagtt ttgagagttt 220

<210> 47  
 <211> 172  
 <212> DNA  
 <213> Homo sapiens

<400> 47  
 acagcctggt catgtaacct gtccttctcc cagccatgcc caccctaacc ctttttccat 60  
 ttctttacgc ttcagaggag cctgcagggt ctgtcaagcc ttcgaaagcc tcagactgta 120  
 cgttgctgtc accttgggga caaccagggg agtggggcct tgggttttgg ct 172

<210> 48  
 <211> 200  
 <212> DNA  
 <213> Homo sapiens

<400> 48  
 ccgaccctc tgattgccac ttgtgtctcc cagtggatga ctacctgggc ttcctcccca 60  
 cttttggggc ctgtacatc aacctctatg gcagtcccag agagttcaca ggcttcccag 120  
 accctacac agagctcaac acaggcaagg taagccggct ggagccctgg caagggcagg 180  
 atgccacatg cccagggtgg 200

<210> 49  
 <211> 217  
 <212> DNA  
 <213> Homo sapiens

<400> 49  
 cctccccctc gtctccccgt ctcttctgtga cctgacctcc ctggcagggg gaaggtgtgg 60  
 cttatcgtgg ccggcttctg ctctccctgg agaccaagct ggtggagcac agtgaacaga 120  
 aggtggagga ccttccctgcg gatgacatcc tccgggtgga ggtgaggggt gtggctctgg 180

37/68

gtgggagctg ggcgtcgagg cagggaagg atggcca

217

<210> 50  
 <211> 269  
 <212> DNA  
 <213> Homo sapiens

<400> 50  
 agcctgggtg cttttctttg ctccctccgt gaccctcttg tctactctct gctctcagaa 60  
 gtaccttagg aggcgcaagt actccctgtt tgcggccttc tactcagcca ccatgctgca 120  
 ggatgtggat gatgccatcc agtttgaggt cagcatcggg aactacggga acaagttcga 180  
 catgacctgc ctgccgctgg cctccaccac tcagtacagc cgtgcagtct ttgacggtga 240  
 ggcagtgtc ctggctggga ccccgatca 269

<210> 51  
 <211> 225  
 <212> DNA  
 <213> Homo sapiens

<400> 51  
 actcctggca cagcgtcag gccgtctct ccattccagg gtgccactac tactacctac 60  
 cctggggtaa cgtgaaacct gtggtggtgc tgtcatccta ctgggaggac atcagccata 120  
 gaatcgagac tcagaaccag ctgcttgagg ttgctgaccg gctggtgagt gaaaacttgc 180  
 ccaaagctgc acatgcctat gcatgcacct gctacccccg ctgca 225

<210> 52  
 <211> 227  
 <212> DNA  
 <213> Homo sapiens

<400> 52  
 ggggtccagca tgcaccctct gccctgtggt gacacacctg acccttgcct gccattcca 60  
 caggaagctg gcctggagca ggtccacctg gccctgaagg cgcagtgtc cacggaggac 120  
 gtggactgc tgggtggtca gctgacggat gagctcatcg caggctgcag gtagggggga 180  
 cctggcgccc ctggtgccc cctctcctgg ctcaactggg cctgttt 227

<210> 53  
 <211> 303  
 <212> DNA  
 <213> Homo sapiens

<400> 53  
 tgggagaccc tgggtcctc aggcgcattc catctgtccg tccctcacag ccagcctctg 60  
 ggtgacatcc atgagacacc ctctgccacc cacctggacc agtacctgta ccagctgcgc 120  
 acccatcacc tgagccaaat cactgaggct gccctggccc tgaagctcgg ccacagtgcg 180  
 ctccctgcag ctctggagca ggcggaggac tggctcctgc gtctgcgtgc cctggcagag 240  
 gaggtaatta agcctggggg tgcctttctt cttctgctct cctgctgcct ggaacatcag 300  
 aac 303

<210> 54  
 <211> 272  
 <212> DNA  
 <213> Homo sapiens

<400> 54  
 cgtgggcctg gtgtgtcacc atccccacc cgaccaccac cctctgttca gccccagaac 60  
 agcctgccgg acatcgatcat ctggatgctg cagggagaca agcgtgtggc ataccagcgg 120  
 gtgcccgcgc accaagtcct cttctcccgg cggggtgcca actactgtgg caagaattgt 180  
 gggaagctac agacaatctt tctgaaagt agttttcttt ttccaagtca tgatcgtatt 240  
 tccaacataa ggcctttctc ccctctctgt ct 272

<210> 55  
 <211> 219  
 <212> DNA  
 <213> Homo sapiens

38/68

<400> 55  
 tgtgggtttc tgtccttctt cggtagccag tatccgatgg agaaggtgcc tggcgccccg 60  
 atgccagtgc agatacgggt caagctgtgg tttgggctct ctgtggatga gaaggagtgc 120  
 aaccagtttg ctgaggggaa gctgtctgtc tttgctgaaa ccgtgagtag ctgccagccc 180  
 ccacctctgc ctccactac ctggagctgc cttggtccc 219

<210> 56  
 <211> 292  
 <212> DNA  
 <213> Homo sapiens

<400> 56  
 tgcctccac tacctggagc tgccttgccc cccttcacgc ctcattcttc ctggccctcc 60  
 agtatgagaa cgagactaag ttggcccttg ttgggaactg ggcacaacg ggccctcacct 120  
 accccaagtt ttctgacgtc acgggcaaga tcaagctacc caaggacagc ttccgccccct 180  
 cggccggctg gacctgggct ggagattggg tcgtgtgtcc ggagaagacg tgagtcgtgg 240  
 gcagggaggg ctggggagag ccaggccagg ctgcccacca tggactgcac cc 292

<210> 57  
 <211> 242  
 <212> DNA  
 <213> Homo sapiens

<400> 57  
 tggatggggg cctctccagc agagcagcag agactctgac cagccctcct ccacagtctg 60  
 ctccatgaca tggacgccgg tcacctgagc ttctgtgaag aggtgtttga gaaccagacc 120  
 cggcttcccc gagggcagtg gatctacatg agtgacaact acaccgatgt ggtaaagcag 180  
 gcactcaggg gcaggtgggg tctagacatt tgggtctctg aggcacctgg tgctcaggga 240  
 ca 242

<210> 58  
 <211> 215  
 <212> DNA  
 <213> Homo sapiens

<400> 58  
 tcacatctgt ctgtctcttc tcattgcttg cctgttcggt tttgtcctta gaacggggag 60  
 aaggtgcttc ccaaggatga cattgagtgc ccactgggct ggaagtggga agatgaggaa 120  
 tggctccacag acctcaaccg ggctgtcgat gagcaaggtg ggcagcatgt ggaacctggc 180  
 gagccccatc cccggcaagc tctcaagcca tgcac 215

<210> 59  
 <211> 246  
 <212> DNA  
 <213> Homo sapiens

<400> 59  
 agagatggtc ccaggagaga tggggggaag tgccaagcaa tgagtgaccg gttccccctc 60  
 ccccaggctg ggagtatagc atcaccatcc ccccggagcg gaagccgaag cactgggtcc 120  
 ctgctgagaa gatgtactac acacaccgac ggcggcgctg ggtgcgcctg cgcaggaggg 180  
 atctcagcca aatggaagca ctgaaaaagg gtgagccagc aggtgggtggg tgggagttag 240  
 gcctgt 246

<210> 60  
 <211> 253  
 <212> DNA  
 <213> Homo sapiens

<400> 60  
 cttccaccg gcctctgagt ctgccccttc ttgtgcagca caggcaggcg gaggcggagg 60  
 gcgagggtct ggagtacgcc tctctttttg gctggaagtt ccacctcgag taccgcaaga 120  
 cagatgcctt ccgcgcgcgc cgctggcgcc gtcgcatgga gccactggag aagacggggc 180  
 ctgcagctgt gtttgccctt gagggggccc tggatgtggt ggctgcactt gtcctggctt 240  
 gggtagggta tat 253

<210> 61  
 <211> 177



39/68

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 61

|            |            |            |            |             |            |     |
|------------|------------|------------|------------|-------------|------------|-----|
| gaatctgcca | taaccagctt | cgtgtctcca | gggcggcgtg | atggatgaca  | agagtgaaga | 60  |
| ttccatgtcc | gtctccacct | tgagcttcgg | tgtgaacaga | cccacgattt  | cctgcatatt | 120 |
| cgactgtaag | taggcttcga | ggcctctatg | gggtgataag | gggtgtgtcac | cttatgc    | 177 |

&lt;210&gt; 62

&lt;211&gt; 181

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 62

|            |            |            |            |             |            |     |
|------------|------------|------------|------------|-------------|------------|-----|
| aaccactcca | gccactcact | ctggcacctc | tgttttttcc | cttgggtgaag | atgggaaccg | 60  |
| ctaccatcta | cgctgctaca | tgtaccaggc | ccgggacctg | gctgcatggg  | acaaggactc | 120 |
| tttttctggg | aggtgggaga | gaggcaggag | agtcagagac | tgtgggctga  | gatctgggaa | 180 |
| t          |            |            |            |             |            | 181 |

&lt;210&gt; 63

&lt;211&gt; 319

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 63

|            |            |            |            |            |            |     |
|------------|------------|------------|------------|------------|------------|-----|
| ccccacatgg | ctctggagaa | gacatctctc | agggtccctg | ctgtgtaatg | tctccctcc  | 60  |
| ccctctggcc | atgcagatcc | ctatgccatc | gtctccttcc | tgcaccagag | ccagaagacg | 120 |
| gtgggtggta | agaacaccct | taacccacc  | tgggaccaga | cgctcatctt | ctacgagatc | 180 |
| gagatctttg | gcgagccggc | cacagttgct | gagcaaccgc | ccagcattgt | ggaggagctg | 240 |
| tacgaccatg | acacttatgt | gagtcctgcc | agtcctgcc  | tcgtccctc  | acagggaggg | 300 |
| accatgtgca | aagggtggg  |            |            |            |            | 319 |

&lt;210&gt; 64

&lt;211&gt; 249

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 64

|            |            |            |            |            |            |     |
|------------|------------|------------|------------|------------|------------|-----|
| gccctgggta | agggatgctg | attcttgtct | ctctacgctt | ggctctaggg | gcagacgagt | 60  |
| ttatgggtcg | ctgcatctgt | caaccgagtc | tggaacggat | gccacggctg | gcctgggtcc | 120 |
| cactgacgag | gggcagccag | ccgtcggggg | agctgctggc | ctcttttgag | ctcatccaga | 180 |
| gagagaaggt | gaggctggtc | tatatccaga | tccaggaggc | ccaggcagga | gtgggggtgg | 240 |
| ggccaaccc  |            |            |            |            |            | 249 |

&lt;210&gt; 65

&lt;211&gt; 158

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 65

|            |            |            |            |            |            |     |
|------------|------------|------------|------------|------------|------------|-----|
| cactgacata | gtccatgagt | gtcatgaggg | tgatgggggc | cttaggtgac | aagcacatga | 60  |
| ccagagctct | cttttcttca | ctccagccgg | ccatccacca | tattcctggg | tttgaggtaa | 120 |
| gtcttgctct | gacctttcct | tcttcaaact | gattgcca   |            |            | 158 |

&lt;210&gt; 66

&lt;211&gt; 132

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 66

|             |            |            |             |            |            |     |
|-------------|------------|------------|-------------|------------|------------|-----|
| ctttttcccc  | ttccaacccc | tctcaccatc | tcctgatgtg  | cacatcccat | ggctgtgggc | 60  |
| cagggtgcagg | agacatcaag | gatcctggat | gagggtgagct | ggcggggccc | aggtagaggg | 120 |
| aaggatgaagc | ca         |            |             |            |            | 132 |

&lt;210&gt; 67

&lt;211&gt; 216

&lt;212&gt; DNA

40/68

&lt;213&gt; Homo sapiens

&lt;400&gt; 67

|            |             |             |            |            |            |     |
|------------|-------------|-------------|------------|------------|------------|-----|
| tcttccttcc | acctttgtct  | ccattctacc  | tgctgtccac | tgcagtctga | ggacacagac | 60  |
| ctgccctacc | caccacccca  | gagggaggcc  | aacatctaca | tggttcctca | gaacatcaag | 120 |
| ccagcgctcc | agcgtaccgc  | catcgagggtg | agccgtccgg | gcctgggcgt | gggggctggg | 180 |
| agcagcctgc | ccttccccctt | cctggcccca  | gccttt     |            |            | 216 |

&lt;210&gt; 68

&lt;211&gt; 263

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 68

|             |            |            |            |            |            |     |
|-------------|------------|------------|------------|------------|------------|-----|
| ccccgggcctt | ctgagccact | ctcctcattc | tgtgtgctta | gaatcctggc | atggggcctg | 60  |
| cggaacatga  | agagttacca | gctggccaac | atctcctccc | ccagcctcgt | ggtagagtgt | 120 |
| gggggcccaga | cggtgcagtc | ctgtgtcatc | aggaacctcc | ggaagaacct | caactttgac | 180 |
| atctgcaccc  | tcttcatgga | agtggtgagc | cccacctccc | tactgtcccc | ttccagagtc | 240 |
| ctgggggctag | aagttctaca | tgt        |            |            |            | 263 |

&lt;210&gt; 69

&lt;211&gt; 249

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 69

|            |             |            |            |            |            |     |
|------------|-------------|------------|------------|------------|------------|-----|
| caggccagtg | cgttcttctt  | cctccaccca | gatgctgccc | agggaggagc | tctactgccc | 60  |
| ccccatcacc | gtcaagggtca | tcgataaacg | ccagtttgcc | cgccggcctg | tggtgggcca | 120 |
| gtgtaccatc | cgctccctgg  | agagcttctt | gtgtgacccc | tactcggcgg | agagtccatc | 180 |
| cccacagggt | ggcccaggta  | ggggaagggg | agatgatggg | caggtcaggg | aagggggagc | 240 |
| ctagggcaa  |             |            |            |            |            | 249 |

&lt;210&gt; 70

&lt;211&gt; 180

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 70

|            |            |            |            |            |            |     |
|------------|------------|------------|------------|------------|------------|-----|
| aggggcgagc | cttttgagag | agcccctgtc | aggcctggat | ggctccctcc | cctgcagacg | 60  |
| atgtgagcct | actcagtcct | ggggaagacg | tgctcatcga | cattgatgac | aaggagcccc | 120 |
| tcaccccat  | ccaggttaga | tgggcatect | ccaggagggc | ctgggtcacc | tttccccctc | 180 |

&lt;210&gt; 71

&lt;211&gt; 211

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 71

|            |            |            |            |            |            |     |
|------------|------------|------------|------------|------------|------------|-----|
| tgctgcttgg | cgagtcctgt | ttctgaaatg | gtctctttct | ttctaccac  | tcaggaggaa | 60  |
| gagttcatcg | attgggtggg | caaattcttt | gcctccatag | gggagaggga | aaagtgcggc | 120 |
| tcctacctgg | agaaggattt | tgacaccctg | aaggtaaggc | ctctcttcag | tctgacagtc | 180 |
| ggtgtgtgtg | tgcgtgctgg | gcagtgggag | a          |            |            | 211 |

&lt;210&gt; 72

&lt;211&gt; 235

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 72

|            |            |            |            |            |            |     |
|------------|------------|------------|------------|------------|------------|-----|
| gttctacttt | ctttctgtct | cttgtcccc  | cctctaatac | ccatgtgtgg | caggtctatg | 60  |
| acacacagct | ggagaatgtg | gaggcctttg | agggcctgtc | tgacttttgt | aacaccttca | 120 |
| agctgtaccg | gggcaagacg | caggaggaga | cagaagatcc | atctgtgatt | ggtgaattta | 180 |
| aggtaaatcc | tcgaagacgt | ccctaacc   | ggtgggccta | agactgtggt | gttgg      | 235 |

&lt;210&gt; 73

&lt;211&gt; 268

&lt;212&gt; DNA

41/68

&lt;213&gt; Homo sapiens

&lt;400&gt; 73

|             |            |             |            |             |             |     |
|-------------|------------|-------------|------------|-------------|-------------|-----|
| ggggacacag  | ccaaaccata | tcaacaatga  | tgataaaata | aaattaaccc  | ttcctttcttt | 60  |
| tcaggggcctc | ttcaaaatth | atccccctccc | agaagaccca | gccatcccca  | tgcccccaag  | 120 |
| acagttccac  | cagctggccg | cccagggacc  | ccaggagtgc | ttgggtccgta | tctacattgt  | 180 |
| ccgagcattt  | ggcctgcagc | ccaaggaccc  | caatggaaag | gtaactttct  | agagccctca  | 240 |
| cctccccaga  | gtagcaggct | caggtaca    |            |             |             | 268 |

&lt;210&gt; 74

&lt;211&gt; 200

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 74

|             |            |            |            |             |             |     |
|-------------|------------|------------|------------|-------------|-------------|-----|
| tttggaaggt  | gttttcacag | aagtgttttg | tctcctcctc | cagtgtgatac | cttacatcaa  | 60  |
| gatctccata  | gggaagaaat | cagtgtgtga | ccaggataac | tacatccctt  | gcacgtctgga | 120 |
| gcccgtattt  | ggaaagtaaa | ttggggcctc | ttgggtcttg | gggtggagga  | gccagacagg  | 180 |
| ataaccacaca | gtctagtggg |            |            |             |             | 200 |

&lt;210&gt; 75

&lt;211&gt; 263

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 75

|             |            |             |            |            |            |     |
|-------------|------------|-------------|------------|------------|------------|-----|
| cctgtttccct | tggtgtccct | gtgtttggctg | acattcgagg | atctgcccct | tcctgcagga | 60  |
| tggttcgagct | gacctgcact | ctgcctcttg  | agaaggacct | aaagatcact | ctctatgact | 120 |
| atgacctcct  | ctccaaggac | gaaaagatcg  | gtgagacggg | cgtcgacctg | gagaacaggc | 180 |
| tgctgtccaa  | gtttggggct | cgtgtgtggac | tcccacagac | ctactgtgtg | tacgtggatg | 240 |
| ggggctggct  | gcctgcttct | ctg         |            |            |            | 263 |

&lt;210&gt; 76

&lt;211&gt; 237

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 76

|            |            |            |            |            |            |     |
|------------|------------|------------|------------|------------|------------|-----|
| aagcatctcg | tctatgtctt | gtgcttgctc | ctcagctctg | gaccgaacca | gtggcgggac | 60  |
| cagctccgcc | cctcccagct | cctccacctc | ttctgccagc | agcatagagt | caaggcacct | 120 |
| gtgtaccgga | cagaccgtgt | aatgtttcag | gataaagaat | attccattga | agagataggt | 180 |
| gagctgccac | atgaccccaa | accatgggtg | gctctcgctg | tatccctccc | tctctca    | 237 |

&lt;210&gt; 77

&lt;211&gt; 245

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 77

|            |            |            |            |            |            |     |
|------------|------------|------------|------------|------------|------------|-----|
| tctctcgctt | cccagctcc  | tgcaactttt | ttgtgttctc | tctggggcag | aggctggcag | 60  |
| gatcccaaac | ccacacctgg | gcccagtggg | ggagcgtctg | gctctgcatg | tgcttcagca | 120 |
| gcagggcctg | gtcccggagc | acgtggagtc | acggcccctc | tacagccccc | tgagccaga  | 180 |
| catcgagcag | gtaggacctt | acccttggtc | ccagagtcc  | cgaactccag | aagcccaacc | 240 |
| ccagg      |            |            |            |            |            | 245 |

&lt;210&gt; 78

&lt;211&gt; 214

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 78

|            |            |            |            |            |            |     |
|------------|------------|------------|------------|------------|------------|-----|
| ggtgcttggt | aacagctggg | taaatgagaa | gggtggggag | agaacggacc | tgtctccgca | 60  |
| ggggaagctg | gggaagctgc | agatgtgggt | cgacctattt | ccgaaggccc | tggggcggcc | 120 |
| tggaacctcc | ttcaacatca | ccccacggag | agccagaagg | tgacttccca | gccacaggct | 180 |
| ctgagctggg | ctgaggggtg | gggcgttgca | gcct       |            |            | 214 |

42/68

<210> 79  
 <211> 229  
 <212> DNA  
 <213> Homo sapiens

<400> 79  
 ttcttaaggc cttcccatcc tttggttagga aatctaggtg gattagagtg atacctttcc 60  
 ccagggttttt cctgcgttgt attatctgga ataccagaga tgtgatcctg gatgacctga 120  
 gcctcacggg ggagaagatg agcgacattt atgtgaaagg gtagggagcc agcgctcctct 180  
 tgccctgtcca gcttcccgcga gctcccgtgc tccctctggg ttgtgcaca 229

<210> 80  
 <211> 261  
 <212> DNA  
 <213> Homo sapiens

<400> 80  
 acgatgtata tactgtgttg gaaatcttaa tgagaactat tctctaaaaa catgtatgtc 60  
 tagttggatg attggctttg aagaacacaa gcaaaagaca gacgtgcatt atcgttccct 120  
 gggagggtgaa ggcaacttca actggagggtt cattttcccc ttcgactacc tgccagctga 180  
 gcaagtctgt accattgccca agaaggtcag tgtccttccg attccctgtg gtgccagcac 240  
 cagggtcttct aaagttagcc t 261

<210> 81  
 <211> 234  
 <212> DNA  
 <213> Homo sapiens

<400> 81  
 tgccctctctc taactttgct tccctgcac cttctctgtt cctcttccgg gtcaggatgc 60  
 cttctggagg ctggacaaga ctgagagcaa aatcccagca cgagtgggtg tccagatctg 120  
 ggacaatgac aagtctctct ttgatgattt tctgggtgatt ttctgggtaa gcgctattgc 180  
 tagaatccca ttctgcacat gggggctgcc ccagaaccca cactgtgtgt ttat 234

<210> 82  
 <211> 297  
 <212> DNA  
 <213> Homo sapiens

<400> 82  
 ggctacaggc tggcagtgat cgagaaaccc ggccaaaaac cacctctctg ttgcaggctc 60  
 cctgcagctc gatctcaacc gcatgcccaa gccagccaag acagccaaga agtgctcctt 120  
 ggaccagctg gatgatgctt tccacccaga atggtttgtg tccctttttg agcagaaaaac 180  
 agtgaagggc tgggtggcct gtgtagcaga agagggtgag aagaaaatac tggcggttaag 240  
 tctacttcct ccagccccag tggaggggcat gggggaagct tcttccatag aaattgt 297

<210> 83  
 <211> 237  
 <212> DNA  
 <213> Homo sapiens

<400> 83  
 cctggttact ctccaggcca ctgagcagag ccttcgtgcc cctaaccaag tgcctctctgt 60  
 cccctcaggg caagctggaa atgaccttg agattgtagc agagagtga catgaggagc 120  
 ggctgtctg ccagggccgg gatgagccca acatgaaccc taagcttgag gacccaagggt 180  
 cagtgtcccag cccctgagcc ccaatgccca caggtctggg ggtataggca cagtcca 237

<210> 84  
 <211> 252  
 <212> DNA  
 <213> Homo sapiens

<400> 84  
 ccctagtaaa ggatgccag ttgactccgg gatctcgtt ccaggcgccc cgacacctcc 60  
 ttccctgtgg ttacctcccc atacaagacc atgaagttca tccctgtggc gcgtttccgg 120  
 tgggccatca tctcttcat catcctcttc atcctgtgct tgttccctgg catcttcctc 180  
 tacgccttcc cgggtgagcag gcctgacgac actgtggtgg gggaaactct ggtctaattg 240

43/68

gggagttcat ca

252

<210> 85  
 <211> 391  
 <212> DNA  
 <213> Homo sapiens

<400> 85  
 tggctgtgcc tgccccagtg ggatcacat ggggccctgt ctctccctc cctccagaac 60  
 tatgctgcca tgaagctggg gaagcccttc agctgaggac tctcctgccc tgtagaaggg 120  
 gccgtggggg cccctccagc atgggactgg cctgcctcct ccgcccagct cggcgagctc 180  
 ctccagacct cctaggcctg attgtcctgc caggggtgggc agacagacag atggaccggc 240  
 ccacactccc agagttgcta acatggagct ctgagatcac ccacttcca tcatttcctt 300  
 ctcccccaac ccaacgcttt tttggatcag ctccagacata tttcagtata aaacagttgg 360  
 aaccacaaaa aaaaaaaaaa aaaaaaaaaa a 391

<210> 86  
 <211> 51  
 <212> PRT  
 <213> Homo sapiens

<400> 86  
 Lys Lys Arg Thr Lys Val Ile Lys Asn Ser Val Asn Pro Val Trp Asn  
 1 5 10 15  
 Glu Gly Phe Glu Trp Asp Leu Lys Gly Ile Pro Leu Asp Gln Gly Ser  
 20 25 30  
 Glu Leu His Val Val Val Lys Asp His Glu Thr Met Gly Arg Asn Arg  
 35 40 45  
 Phe Leu Gly  
 50

<210> 87  
 <211> 45  
 <212> PRT  
 <213> Homo sapiens

<400> 87  
 Ser Lys Ile Leu Glu Lys Thr Ala Asn Pro Gln Trp Asn Gln Asn Ile  
 1 5 10 15  
 Thr Leu Pro Ala Met Phe Pro Ser Met Cys Glu Lys Met Arg Ile Arg  
 20 25 30  
 Ile Ile Asp Trp Asp Arg Leu Thr His Asn Asp Ile Val  
 35 40 45

<210> 88  
 <211> 82  
 <212> PRT  
 <213> Homo sapiens

<400> 88  
 Gln Ala Arg Asp Leu Ala Ala Met Asp Lys Asp Ser Phe Ser Asp Pro  
 1 5 10 15  
 Tyr Ala Ile Val Ser Phe Leu His Gln Ser Gln Lys Thr Val Val Val  
 20 25 30  
 Lys Asn Thr Leu Asn Pro Thr Trp Asp Gln Thr Leu Ile Phe Tyr Glu  
 35 40 45  
 Ile Glu Ile Phe Gly Glu Pro Ala Thr Val Ala Glu Gln Pro Pro Ser  
 50 55 60  
 Ile Val Val Glu Leu Tyr Asp His Asp Thr Tyr Gly Ala Asp Glu Phe  
 65 70 75 80  
 Met Gly

<210> 89  
 <211> 79  
 <212> PRT  
 <213> Homo sapiens

44/68

<400> 89  
 Ile Tyr Ile Val Arg Ala Phe Gly Leu Gln Pro Lys Asp Pro Asn Gly  
   1                  5                  10                  15  
 Lys Cys Asp Pro Tyr Ile Lys Ile Ser Ile Gly Lys Lys Ser Val Ser  
                   20                  25                  30  
 Asp Gln Asp Asn Tyr Ile Pro Cys Thr Leu Glu Pro Val Phe Gly Lys  
                   35                  40                  45  
 Met Phe Glu Leu Thr Cys Thr Leu Pro Leu Glu Lys Asp Leu Lys Ile  
                   50                  55                  60  
 Thr Leu Tyr Asp Tyr Asp Leu Leu Ser Lys Asp Glu Lys Ile Gly  
   65                  70                  75

<210> 90  
 <211> 152  
 <212> DNA  
 <213> Homo sapiens

<400> 90  
 acgatgtata tactgtgttg gaaatcttaa tgagaactat tctctaaaaa catgtatgtc 60  
 tagttggatg attggctttg aagaacacaa gcaaaagaca gacgtgcatt atcgttccct 120  
 gggaggtgaa ggcaacttca actggagggt ca 152

<210> 91  
 <211> 56  
 <212> DNA  
 <213> Homo sapiens

<400> 91  
 gtcagtgtcc ttccgattcc ctgtggtgcc agcaccaggg cttctaaagt tagcct 56

<210> 92  
 <211> 55  
 <212> DNA  
 <213> Homo sapiens

<400> 92  
 tgccctcttc taacttttgc tctttgcac cttctctgtt cctcttccgg gtcag 55

<210> 93  
 <211> 68  
 <212> DNA  
 <213> Homo sapiens

<400> 93  
 gtaagcgcta ttgctagaat cccattctgc acatgggggc tgccccagaa cccacactgt 60  
 gtgtttat 68

<210> 94  
 <211> 56  
 <212> DNA  
 <213> Homo sapiens

<400> 94  
 ggctacaggc tggcagtgat cgagaaaccc ggccaaaaac cacctctctg ttgcag 56

<210> 95  
 <211> 62  
 <212> DNA  
 <213> Homo sapiens

<400> 95  
 gtaagtctac ttcctccagc cccagtggag ggcattgggg aagcttcttc catagaaatt 60  
 gt 62

<210> 96  
 <211> 68  
 <212> DNA

45/68

<213> Homo sapiens  
 <400> 96  
 cctgggttact ctccaggcca ctgagcagag ccttcgtgcc cctaaccaag tgctctctgt 60  
 cccctcag 68  
 <210> 97  
 <211> 59  
 <212> DNA  
 <213> Homo sapiens  
 <400> 97  
 gtcagtgtccc agcccctgag ccccaatgcc cacaggtctg ggggtatagg cacagtcca 59  
 <210> 98  
 <211> 44  
 <212> DNA  
 <213> Homo sapiens  
 <400> 98  
 ccctagtaaa ggatgcccag ttgactccgg gatctcgctt ccag 44  
 <210> 99  
 <211> 60  
 <212> DNA  
 <213> Homo sapiens  
 <400> 99  
 gtgagcaggc ctgacgacac tgtggtgggg gaactctggg tctaattgggg gagttcatca 60  
 <210> 100  
 <211> 57  
 <212> DNA  
 <213> Homo sapiens  
 <400> 100  
 tggctgtgcc tgccccagtg ggatcaccat gggtcctgt ctctccctc cctccag 57  
 <210> 101  
 <211> 23  
 <212> DNA  
 <213> Homo sapiens  
 <400> 101  
 tctcttctcc tagaggcca tag 23  
 <210> 102  
 <211> 24  
 <212> DNA  
 <213> Homo sapiens  
 <400> 102  
 ctgttctctc ccatcgcttc atgg 24  
 <210> 103  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens  
 <400> 103  
 gctcctccc tgaccctctg 20  
 <210> 104  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens

|  |    |
|--|----|
| <400> 104<br>gggtcccagc caggagcact g                     | 21 |
| <210> 105<br><211> 24<br><212> DNA<br><213> Homo sapiens |    |
| <400> 105<br>cccctctcac catctcctga tgtg                  | 24 |
| <210> 106<br><211> 25<br><212> DNA<br><213> Homo sapiens |    |
| <400> 106<br>tggcttcacc ttccctctac ctgg                  | 25 |
| <210> 107<br><211> 24<br><212> DNA<br><213> Homo sapiens |    |
| <400> 107<br>tcctttggtg ggaatctag gtgg                   | 24 |
| <210> 108<br><211> 21<br><212> DNA<br><213> Homo sapiens |    |
| <400> 108<br>ggaagctgga caggcaagag g                     | 21 |
| <210> 109<br><211> 27<br><212> DNA<br><213> Homo sapiens |    |
| <400> 109<br>atatactgtg ttggaaatct taatgag               | 27 |
| <210> 110<br><211> 21<br><212> DNA<br><213> Homo sapiens |    |
| <400> 110<br>gctggcacca cagggaatcg g                     | 21 |
| <210> 111<br><211> 25<br><212> DNA<br><213> Homo sapiens |    |
| <400> 111<br>ctttgcttcc ttgcatcctt ctctg                 | 25 |
| <210> 112<br><211> 21<br><212> DNA<br><213> Homo sapiens |    |
| <400> 112<br>agcccccattg tgcagaatgg g                    | 21 |



47/68

<210> 113  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 113  
ggcagtgatc gagaaacccg g 21

<210> 114  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 114  
catgccctcc actggggctg g 21

<210> 115  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 115  
ggatgcccag ttgactccg g 21

<210> 116  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 116  
ccccaccaca gtgtcgtcag g 21

<210> 117  
<211> 6240  
<212> DNA  
<213> Homo sapiens

<400> 117

|             |           |           |           |          |           |      |
|-------------|-----------|-----------|-----------|----------|-----------|------|
| atgctgagg   | tcttcac   | ctatgcc   | aacgtcc   | cacccga  | cgacatc   | 60   |
| gatgcctac   | gtccgcg   | gtttgcag  | gtgaaga   | gaaccaa  | catcaaga  | 120  |
| agcgtgaac   | ctgtatg   | tgagggat  | gaatggg   | tcaaggg  | ccccctg   | 180  |
| cagggctct   | agcttcat  | ggtggtcaa | gacatgag  | cgatggg  | gaacagg   | 240  |
| ctgggggaa   | ccaaggtcc | actccgag  | gtcctcg   | cccctagt | gtccgcc   | 300  |
| ttcaatgccc  | ccctgctg  | caccaaga  | cagcccac  | gggcctcg | ggtcctg   | 360  |
| gtgtcctaca  | caccgtgcc | tggagctgt | cccctgttc | cgccccct | tcctctg   | 420  |
| ccctccccga  | ctctgcct  | cctggatgt | gtggcag   | caggagg  | ggaagac   | 480  |
| gaggaccagg  | gactcact  | agatgagg  | gagccatt  | tggatcaa | cggaggccc | 540  |
| ggggctccca  | ccaccccaa | gaaactac  | tcacgtc   | cgccccac | ccccggg   | 600  |
| aaaagaaagc  | gaagtgcg  | tacatcta  | aagctgct  | cagacaa  | gcaggatt  | 660  |
| cagatcagg   | tccaggtg  | cgaggggc  | cagctgcc  | gggtgaac | caagcctg  | 720  |
| gtcaagggtta | ccgctgcag | gcagacca  | cggacgcg  | tccacaag | aaacagccc | 780  |
| ctcttcaatg  | agactcttt | cttcaact  | tttgactc  | ctggggag | gtttgatg  | 840  |
| cccatcttta  | tcacggtg  | agactctc  | tctctcag  | cagatgct | cctcgggg  | 900  |
| ttccggatgg  | acgtgggc  | catttac   | gagccccg  | acgcctat | caggaagt  | 960  |
| ctgctgctct  | cagaccct  | tgacttct  | gctggggc  | gaggctac | gaaaaca   | 1020 |
| ctttgtgtgc  | tggggcct  | ggacgaag  | cctctgg   | gaaaagac | ctctgaag  | 1080 |
| aaggaggaca  | ttgaaagc  | cctgctcc  | cccacagg  | tagccctg | aggagccc  | 1140 |
| ttctgcctga  | aggtcttc  | ggccgagg  | ttgccgc   | tggacgat | cgtgatgg  | 1200 |
| aacgtgaaac  | agatcttt  | cctcgag   | aacaaga   | acttggtg | cccctttg  | 1260 |
| gaggtcagct  | ttgcgggg  | aatgctgt  | agcaagat  | tggagaag | ggccaacc  | 1320 |
| cagtggaaac  | agaacatc  | actgcctg  | atgtttcc  | ccatgtgc | aaaaatg   | 1380 |
| attcgtatca  | tagactgg  | ccgcctga  | cacaatga  | tcgtggct | cacctacc  | 1440 |
| agtatgtcga  | aaatctct  | ccctggag  | gaaataga  | aggagcct | aggtgctg  | 1500 |
| aagccttcga  | aagcctca  | cctggatg  | tacctggg  | tcctcccc | ttttgggc  | 1560 |
| tgctacatca  | accttatg  | cagtccc   | gagttcac  | gcttccc  | cccctaca  | 1620 |
| gagctcaaca  | caggcaag  | ggaaggtg  | gcttatcg  | gccggctt | gctctccc  | 1680 |
| gagaccaagc  | tggtggag  | cagtgaac  | aaggtgg   | accttcct | ggatgac   | 1740 |

|             |             |            |             |             |            |      |
|-------------|-------------|------------|-------------|-------------|------------|------|
| ctccgggtgg  | agaagtacct  | taggaggcgc | aagtactccc  | tgtttgcggc  | cttctactca | 1800 |
| gccaccatgc  | tgacagatgt  | ggatgatgcc | atccagtttg  | aggtcagcat  | cggaactac  | 1860 |
| gggaacaagt  | tcgacatgac  | ctgcctgccg | ctggcctcca  | ccactcagta  | cagccgtgca | 1920 |
| gtctttgacg  | ggtgccacta  | ctactacctt | ccctggggta  | acgtgaaacc  | tgtggtggtg | 1980 |
| ctgtcatcct  | actgggagga  | catcagccat | agaatcgaga  | ctcagaacca  | gctgcttggg | 2040 |
| attgctgacc  | ggctggaagc  | tggcctggag | caggtccacc  | tggccctgaa  | ggcgagtg   | 2100 |
| tccacggagg  | acgtggactc  | gctgggtgct | cagctgacgg  | atgagctcat  | cgagcgctgc | 2160 |
| agccagcctc  | tgggtgacat  | ccatgagaca | ccctctgcca  | cccacctgga  | ccagtacctg | 2220 |
| taccagctgc  | gcacccatca  | cctgagccaa | atcactgagg  | ctgccctggc  | cctgaagctc | 2280 |
| ggccacagtg  | agctccctgc  | agctctggag | caggcggagg  | actggctcct  | gcgtctgcgt | 2340 |
| gcccctggcg  | aggagcccca  | gaacagcctg | ccggacatcg  | tcactctggat | gctgcaggga | 2400 |
| gacaagcgctg | tggcatacca  | gcgggtgccc | gcccaccaag  | tcctcttctc  | ccggcggggt | 2460 |
| gccaactact  | gtggcaagaa  | ttgtgggaag | ctacagacaa  | tctttctgaa  | atatccgatg | 2520 |
| gagaaggtgc  | ctggcgcccc  | gatgccagtg | cagatacggg  | tcaagctgtg  | gtttgggctc | 2580 |
| tctgtggatg  | agaaggagtt  | caaccagttt | gctgagggga  | agctgtctgt  | ctttgctgaa | 2640 |
| acctatgaga  | acgagactaa  | gttggccctt | gttgggaact  | ggggcacaa   | ggcctcacc  | 2700 |
| taccccaagt  | tttctgacgt  | cacgggcaag | atcaagctac  | ccaaggacag  | cttccgcccc | 2760 |
| tcggccggct  | ggacctgggc  | tggagattgg | ttcgtgtgtc  | cggagaagac  | tctgctccat | 2820 |
| gacatggacg  | ccggtcacct  | gagcttcgtg | gaagaggtgt  | ttgagaacca  | gacccggctt | 2880 |
| cccggaggcc  | agtggatcta  | catgagtgc  | aactacaccg  | atgtgaacgg  | ggagaaggtg | 2940 |
| cttcccaagg  | atgacattga  | gtgccactg  | ggctggaagt  | gggaagatga  | ggaatggctc | 3000 |
| acagacctca  | accgggctgt  | cgatgagcaa | ggctgggagt  | atagcatcac  | catccccccg | 3060 |
| gagcgggaagc | cgaagcactg  | ggtccctgct | gagaagatgt  | actacacaca  | ccgacggcgg | 3120 |
| cgctgggtgc  | gcctgcgcag  | gagggatctc | agccaaatgg  | aagcactgaa  | aaggcacagg | 3180 |
| caggcggagg  | cggaggcgca  | gggtctggag | tacgcctctc  | tttttggctg  | gaagttccac | 3240 |
| ctcgagtacc  | tgccctccgc  | tgccctccgc | cgcgcgcgt   | ggcgccgtcg  | catggagcca | 3300 |
| ctggagaaga  | cggggcctgc  | agctgtgttt | gcccttgagg  | gggccctggg  | cgccgtgatg | 3360 |
| gatgacaaga  | gtgaagattc  | catgtccgtc | tccaccttga  | gcttcggtgt  | gaacagaccc | 3420 |
| acgatttctt  | gcatattcga  | ctatgggaac | cgctaccatc  | tacgctgcta  | catgtaccag | 3480 |
| gcccgggacc  | tggctgcgat  | ggacaaggac | tctttttctg  | atccctatgc  | catcgtctcc | 3540 |
| ttcctgcacc  | agagccagaa  | gacggtgggt | gtgaagaaca  | cccttaaccc  | cacctgggac | 3600 |
| cagacgctca  | tcttctacga  | gatcgagatc | tttggcgagc  | cggccacagt  | tgctgagcaa | 3660 |
| ccgcccagca  | ttgtggtgga  | gctgtacgac | catgacactt  | atggtgcaga  | cgagtttatg | 3720 |
| ggctcgctgca | tctgtcaacc  | gagtctggaa | cggatgccac  | ggctggcctg  | gttcccactg | 3780 |
| acgaggggca  | gccagccgtc  | gggggagctg | ttggcctctt  | ttgagctcat  | ccagagagag | 3840 |
| aagccggcca  | tccaccatat  | tccctggttt | gaggtgcagg  | agacatcaag  | gatcctggat | 3900 |
| gagctctgag  | acacagacct  | gccctaccca | ccaccccaaga | gggaggccaa  | catctacatg | 3960 |
| gttccctcaga | acatcaagcc  | agcgtccag  | cgtaccgcca  | tcgagatcct  | ggcatggggc | 4020 |
| ctgcggaaca  | tgaagagtta  | ccagctggcc | aacatctcct  | ccccagcct   | cgtggtagag | 4080 |
| tgtgggggccc | agacggtgca  | gtcctgtgtc | atcaggaaac  | tccggaagaa  | ccccaacttt | 4140 |
| gacatctgca  | ccctcttcac  | ggaagtgatg | ctgcccaggg  | aggagctcta  | ctgccccccc | 4200 |
| atcacctgca  | aggtcatcga  | taaccgccag | tttggccgcc  | ggcctgtggt  | gggccagtgt | 4260 |
| accatccgct  | ccctggagag  | cttccctgtg | gacccctact  | cggcggagag  | tccatcccc  | 4320 |
| caggggtggcc | cagacgatgt  | gagcctactc | agtcctgggg  | aagacgtgct  | catcgacatt | 4380 |
| gatgacaagg  | agccctcat   | ccccatccag | gaggaagagt  | tcacgcattg  | gtggagcaaa | 4440 |
| ttctttgcct  | ccatagggga  | gagggaaaag | tgccgctcct  | acctggagaa  | ggattttgac | 4500 |
| accctgaagg  | tctatgacac  | acagctggag | aatgtggagg  | cctttgaggg  | cctgtctgac | 4560 |
| ttttgttaaca | ccttcaagct  | gtaccggggc | aagacgcagg  | aggagacaga  | agatccatct | 4620 |
| gtgattgggtg | aatttaaggg  | cctcttcaaa | atttatcccc  | tcccagaaga  | ccagccatc  | 4680 |
| cccatgccccc | caagacagtt  | ccaccagctg | gcccggcagg  | gaccccagga  | gtgcttggtc | 4740 |
| cgtatctaca  | ttgtccgagc  | atttggcctg | cagcccaagg  | accccaatgg  | aaagtgtgat | 4800 |
| ccttacatca  | agatctccat  | agggaagaaa | tcagtgagtg  | accaggataa  | ctacatcccc | 4860 |
| tgacacgctg  | agcccgctatt | tggaaagatg | ttcgagctga  | cctgcactct  | gcctctggag | 4920 |
| aaggacctaa  | agatcactct  | ctatgactat | gacctcctct  | ccaaggacga  | aaagatcggt | 4980 |
| gagacgctcg  | tcgacctgga  | gaacaggctg | ctgtcccaagt | ttggggctcg  | ctgtggactc | 5040 |
| ccacagacct  | actgtgtctc  | tggaccgaac | cagtgccggg  | accagctccg  | ccccctccag | 5100 |
| ctcctccacc  | tcttctgcca  | gcagcataga | gtcaaggcac  | ctgtgtaccg  | gacagaccgt | 5160 |
| gtaatgtttc  | aggataaaga  | atattccatt | gaagagatag  | aggctggcag  | gatcccaaac | 5220 |
| ccacacctgg  | gcccagtgga  | ggagcgtctg | gctctgcatt  | tgcttcagca  | gcagggctcg | 5280 |
| gtcccggagc  | acgtggagtc  | acggccctc  | tacagccccc  | tgagccaga   | catcgagcag | 5340 |
| gggaagctgc  | agatgtgggt  | cgacctattt | ccgaaggccc  | tggggcgggc  | tggacctccc | 5400 |
| ttcaacatca  | ccccacggag  | agccagaagg | tttttccctg  | gttgtattat  | ctggaatacc | 5460 |
| agagatgtga  | tccctggatga | cctgagcctc | acgggggaga  | agatgagcga  | catttatgtg | 5520 |
| aaaggttgga  | tgattggctt  | tgaagaacac | aagcaaaaga  | cagacgtgca  | ttatcgctcc | 5580 |
| ctgggaggtg  | aaggcaactt  | caactggagg | ttcattttcc  | ccttcgacta  | cctgccagct | 5640 |
| gagcaagtct  | gtaccattgc  | caagaaggat | gccttctgga  | ggctggacaa  | gactgagagc | 5700 |
| aaaatcccag  | cacgagtgtt  | gttccagatc | tgggacaatg  | acaagttctc  | ctttgatgat | 5760 |

49/68

|            |            |            |            |             |            |      |
|------------|------------|------------|------------|-------------|------------|------|
| tttctgggct | ccctgcagct | cgatctcaac | cgcatgccca | agccagccaa  | gacagccaag | 5820 |
| aagtgtcct  | tgaccagct  | ggatgatgct | ttccacccag | aatggtttgt  | gtcccttttt | 5880 |
| gagcagaaaa | cagtgaagg  | ctggtggccc | tgtgtagcag | aagaggggtga | gaagaaaata | 5940 |
| ctggcgggca | agctggaaat | gaccttgag  | attgtagcag | agagtgcagca | tgaggagcgg | 6000 |
| cctgtggcc  | agggccggga | tgagcccaac | atgaacccta | agcttgagga  | cccaaggcgc | 6060 |
| cccgcacct  | ccttcctgtg | gtttacctcc | ccatacaaga | ccatgaagtt  | catcctgtgg | 6120 |
| cggcgtttcc | ggtgggcat  | catcctcttc | atcatcctct | tcatacctgct | gctgttcctg | 6180 |
| gccatcttca | tctacgcctt | cccgaactat | gctgccatga | agctgggtgaa | gcccttcagc | 6240 |

<210> 118  
 <211> 13  
 <212> DNA  
 <213> Homo sapiens

<400> 118  
 cgcaagcatg ctg 13

<210> 119  
 <211> 12  
 <212> DNA  
 <213> Homo sapiens

<400> 119  
 gagacgatgg gg 12

<210> 120  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens

<400> 120  
 gatctaacc tgctgtcac c 21

<210> 121  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens

<400> 121  
 ctggtgtgtt gcagagcgt g 21

<210> 122  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens

<400> 122  
 cctctcttct gctgtcttca g 21

<210> 123  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens

<400> 123  
 tgtgtctggt tcacctcgt g 21

<210> 124  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens

<400> 124  
 tccaaataga aatgcctgaa c 21

<210> 125  
 <211> 21

50/68

<212> DNA  
<213> Homo sapiens  
  
<400> 125  
aggtatcacc tccaagtgtt g 21  
  
<210> 126  
<211> 21  
<212> DNA  
<213> Homo sapiens  
  
<400> 126  
taccagcttc agagctccct g 21  
  
<210> 127  
<211> 19  
<212> DNA  
<213> Homo sapiens  
  
<400> 127  
ttgatcaggg tgctcttgg 19  
  
<210> 128  
<211> 20  
<212> DNA  
<213> Homo sapiens  
  
<400> 128  
ggagaattgc ttgaaccag 20  
  
<210> 129  
<211> 22  
<212> DNA  
<213> Homo sapiens  
  
<400> 129  
tggctaataga tggtaaacat tt 22  
  
<210> 130  
<211> 21  
<212> DNA  
<213> Homo sapiens  
  
<400> 130  
gaccacaaag cggcgctcg g 21  
  
<210> 131  
<211> 21  
<212> DNA  
<213> Homo sapiens  
  
<400> 131  
gaccccggcg aggggtggctg g 21  
  
<210> 132  
<211> 24  
<212> DNA  
<213> Homo sapiens  
  
<400> 132  
tgtctctcca ttctcccttt tgtg 24  
  
<210> 133  
<211> 24  
<212> DNA  
<213> Homo sapiens

51/68

<400> 133  
aggacactgc tgagaaggca cctc 24

<210> 134  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 134  
agtgccctgg tggcacgaag g 21

<210> 135  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 135  
cctacctgca ccttcaagcc atgg 24

<210> 136  
<211> 23  
<212> DNA  
<213> Homo sapiens

<400> 136  
cagaagagcc aggggtgcctt agg 23

<210> 137  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 137  
ccttggacct taacctggca gagg 24

<210> 138  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 138  
cgaggccagc gcaccaacct g 21

<210> 139  
<211> 22  
<212> DNA  
<213> Homo sapiens

<400> 139  
actgccggcc attcttgctg gg 22

<210> 140  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 140  
ccaggcctca ttagggccct c 21

<210> 141  
<211> 22  
<212> DNA  
<213> Homo sapiens

<400> 141  
ctgaagagga gcctggggtc ag 22

52/68

<210> 142  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 142  
ctgagatttc tgactcttgg ggtg 24

<210> 143  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 143  
aaggttctgc cctcatgccc catg 24

<210> 144  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 144  
ctggcctgag ggatcagcag g 21

<210> 145  
<211> 23  
<212> DNA  
<213> Homo sapiens

<400> 145  
gtgcatacat acagcccacg gag 23

<210> 146  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 146  
gagctattgg gttggccgtg tggg 24

<210> 147  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 147  
accaacacgg agaagtgaga actg 24

<210> 148  
<211> 26  
<212> DNA  
<213> Homo sapiens

<400> 148  
ccacacttta tttaacgctt tggcgg 26

<210> 149  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 149  
cagaaccaa atgcaaggat acgg 24

<210> 150  
<211> 25  
<212> DNA

53/68

<213> Homo sapiens

<400> 150 25  
cttctgattc tgggatcacc aaagg

<210> 151  
<211> 22  
<212> DNA  
<213> Homo sapiens

<400> 151 22  
ggaccgtaag gaagacccag gg

<210> 152  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 152 24  
cctgtgctca ggagcgcag aagg

<210> 153  
<211> 22  
<212> DNA  
<213> Homo sapiens

<400> 153 22  
gcagacctcc cacccaagg cg

<210> 154  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 154 24  
gagacagatg ggggacagtc aggg

<210> 155  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 155 21  
cctcccgaga gaacctcct g

<210> 156  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 156 21  
gggagcccag agtcccatg g

<210> 157  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 157 21  
gggcctcctt gggtttgctg g

<210> 158  
<211> 21  
<212> DNA  
<213> Homo sapiens

|  |    |
|--|----|
| <400> 158<br>gcctccccag catcctgccg g                     | 21 |
| <210> 159<br><211> 24<br><212> DNA<br><213> Homo sapiens |    |
| <400> 159<br>tcactgagcc gaatgaaact gagg                  | 24 |
| <210> 160<br><211> 24<br><212> DNA<br><213> Homo sapiens |    |
| <400> 160<br>tgtggcctga gttcctttcc tgtg                  | 24 |
| <210> 161<br><211> 24<br><212> DNA<br><213> Homo sapiens |    |
| <400> 161<br>ggtcaaaggg cagaacgaag aggg                  | 24 |
| <210> 162<br><211> 21<br><212> DNA<br><213> Homo sapiens |    |
| <400> 162<br>cccgtccttc tcccagccat g                     | 21 |
| <210> 163<br><211> 21<br><212> DNA<br><213> Homo sapiens |    |
| <400> 163<br>ctcccctggg tgtccccaag g                     | 21 |
| <210> 164<br><211> 24<br><212> DNA<br><213> Homo sapiens |    |
| <400> 164<br>cgaccctct gattgccact tgtg                   | 24 |
| <210> 165<br><211> 21<br><212> DNA<br><213> Homo sapiens |    |
| <400> 165<br>ggcatcctgc ccttgccagg g                     | 21 |
| <210> 166<br><211> 20<br><212> DNA<br><213> Homo sapiens |    |
| <400> 166<br>tctgtctccc ctgctccttg                       | 20 |



55/68

<210> 167  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 167  
cttccctgcc ccgacgccca g 21

<210> 168  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 168  
cagcgctcag gcccgctctc c 21

<210> 169  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 169  
tgcataaggca tgtgcagctt tggg 24

<210> 170  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 170  
catgcaccct ctgccctgtg g 21

<210> 171  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 171  
agttgagcca ggagaggtgg g 21

<210> 172  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 172  
catcaggcgc attccatctg tccg 24

<210> 173  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 173  
agcaggagag cagaagaaga aagg 24

<210> 174  
<211> 22  
<212> DNA  
<213> Homo sapiens

<400> 174  
gtgtgtcacc atccccaccc cg 22

<210> 175  
<211> 25  
<212> DNA

<213> Homo sapiens  
<400> 175  
caagagatgg gagaaaggcc ttatg 25  
<210> 176  
<211> 23  
<212> DNA  
<213> Homo sapiens  
<400> 176  
ctgggacatc cggatcctga agg 23  
<210> 177  
<211> 22  
<212> DNA  
<213> Homo sapiens  
<400> 177  
tccaggtagt gggaggcaga gg 22  
<210> 178  
<211> 24  
<212> DNA  
<213> Homo sapiens  
<400> 178  
tcccactacc tggagctgcc ttgg 24  
<210> 179  
<211> 21  
<212> DNA  
<213> Homo sapiens  
<400> 179  
ggctctcccc agccctccct g 21  
<210> 180  
<211> 24  
<212> DNA  
<213> Homo sapiens  
<400> 180  
cagagcagca gagactctga ccag 24  
<210> 181  
<211> 21  
<212> DNA  
<213> Homo sapiens  
<400> 181  
tagaccccac ctgcccctga g 21  
<210> 182  
<211> 24  
<212> DNA  
<213> Homo sapiens  
<400> 182  
tcctctcatt gcttgccctgt tcgg 24  
<210> 183  
<211> 21  
<212> DNA  
<213> Homo sapiens

57/68

<400> 183  
ttgagagctt gccggggatg g 21

<210> 184  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 184  
aagtgccaaag caatgagtga ccgg 24

<210> 185  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 185  
ctcactccca cccaccacct g 21

<210> 186  
<211> 21  
<212> DNA  
<213> Homo sapiens

<400> 186  
cccaccggcc tctgagtctg c 21

<210> 187  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 187  
accctaccca agccaggaca agtg 24

<210> 188  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 188  
gaatctgcca taaccagctt cgtg 24

<210> 189  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 189  
tatcacccca tagaggcctc gaag 24

<210> 190  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 190  
cagccactca ctctggcacc tctg 24

<210> 191  
<211> 24  
<212> DNA  
<213> Homo sapiens

<400> 191  
agcccacagt ctctgactct cctg 24

58/68

<210> 192  
<211> 24  
<212> DNA  
<213> Homo sapiens  
  
<400> 192  
acatctctca gggccctgc tgtg 24  
  
<210> 193  
<211> 21  
<212> DNA  
<213> Homo sapiens  
  
<400> 193  
cctgtgaggg gacgaggcag g 21  
  
<210> 194  
<211> 24  
<212> DNA  
<213> Homo sapiens  
  
<400> 194  
gccctgggta agggatgctg attc 24  
  
<210> 195  
<211> 21  
<212> DNA  
<213> Homo sapiens  
  
<400> 195  
cctgcctggg cctcctggat c 21  
  
<210> 196  
<211> 21  
<212> DNA  
<213> Homo sapiens  
  
<400> 196  
gaggggtgatg ggggccttag g 21  
  
<210> 197  
<211> 24  
<212> DNA  
<213> Homo sapiens  
  
<400> 197  
gcaatcagtt tgaagaagga aagg 24  
  
<210> 198  
<211> 24  
<212> DNA  
<213> Homo sapiens  
  
<400> 198  
cacctttgtc tccattctac ctgc 24  
  
<210> 199  
<211> 21  
<212> DNA  
<213> Homo sapiens  
  
<400> 199  
ctcccagccc ccacgcccag g 21  
  
<210> 200  
<211> 24  
<212> DNA

59/68

<213> Homo sapiens  
<400> 200  
ctgagccact ctcctcattc tgtg 24  
<210> 201  
<211> 21  
<212> DNA  
<213> Homo sapiens  
<400> 201  
tggaagggga cagtagggag g 21  
<210> 202  
<211> 22  
<212> DNA  
<213> Homo sapiens  
<400> 202  
ggccagtgcg ttcttctcc tc 22  
<210> 203  
<211> 22  
<212> DNA  
<213> Homo sapiens  
<400> 203  
tccctgacct gccatcatc tc 22  
<210> 204  
<211> 21  
<212> DNA  
<213> Homo sapiens  
<400> 204  
gcccctgtca ggctggatg g 21  
<210> 205  
<211> 21  
<212> DNA  
<213> Homo sapiens  
<400> 205  
tgaccaggc ctccctggag g 21  
<210> 206  
<211> 24  
<212> DNA  
<213> Homo sapiens  
<400> 206  
ctgaaatggt ctctttcttt ctac 24  
<210> 207  
<211> 24  
<212> DNA  
<213> Homo sapiens  
<400> 207  
cacaccgact gtcagactga agag 24  
<210> 208  
<211> 24  
<212> DNA  
<213> Homo sapiens

60/68

|  |    |
|--|----|
| <400> 208<br>ttgtcccctc ctctaattccc catg                 | 24 |
| <210> 209<br><211> 21<br><212> DNA<br><213> Homo sapiens |    |
| <400> 209<br>gggttaggga cgtcttcgag g                     | 21 |
| <210> 210<br><211> 22<br><212> DNA<br><213> Homo sapiens |    |
| <400> 210<br>cagccaaacc atatcaacaa tg                    | 22 |
| <210> 211<br><211> 21<br><212> DNA<br><213> Homo sapiens |    |
| <400> 211<br>ctggggaggt gagggctcta g                     | 21 |
| <210> 212<br><211> 21<br><212> DNA<br><213> Homo sapiens |    |
| <400> 212<br>gaagtgtttt gtctcctcct c                     | 21 |
| <210> 213<br><211> 20<br><212> DNA<br><213> Homo sapiens |    |
| <400> 213<br>gcaggcagcc agcccccatc                       | 20 |
| <210> 214<br><211> 21<br><212> DNA<br><213> Homo sapiens |    |
| <400> 214<br>gggtgccctg tgttggtga c                      | 21 |
| <210> 215<br><211> 20<br><212> DNA<br><213> Homo sapiens |    |
| <400> 215<br>gcaggcagcc agcccccatc                       | 20 |
| <210> 216<br><211> 24<br><212> DNA<br><213> Homo sapiens |    |
| <400> 216<br>ctcgtctatg tcttggtgctt gctc                 | 24 |

61/68

<210> 217  
 <211> 23  
 <212> DNA  
 <213> Homo sapiens  
  
 <400> 217  
 caccatgggtt tgggggtcatg tgg 23  
  
 <210> 218  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens  
  
 <400> 218  
 tctcgcttcc ccagctcctg c 21  
  
 <210> 219  
 <211> 22  
 <212> DNA  
 <213> Homo sapiens  
  
 <400> 219  
 tctggagttc gaggactctg gg 22  
  
 <210> 220  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens  
  
 <400> 220  
 agaaggggtgg ggagagaacg g 21  
  
 <210> 221  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens  
  
 <400> 221  
 cagctcagag cctgtggctg g 21  
  
 <210> 222  
 <211> 24  
 <212> DNA  
 <213> Homo sapiens  
  
 <400> 222  
 aaggccttcc catcctttgg tagg 24  
  
 <210> 223  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens  
  
 <400> 223  
 acaaccaga gggagcacgg g 21  
  
 <210> 224  
 <211> 25  
 <212> DNA  
 <213> Homo sapiens  
  
 <400> 224  
 gttgacgatg tatatactgt gttgg 25  
  
 <210> 225  
 <211> 25  
 <212> DNA

<213> Homo sapiens  
 <400> 225  
 gcctctctct aactttgctt ccttg 25  
 <210> 226  
 <211> 24  
 <212> DNA  
 <213> Homo sapiens  
 <400> 226  
 ggctacaggc tggcagtgat cgag 24  
 <210> 227  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens  
 <400> 227  
 ttcccccatg ccctccactg g 21  
 <210> 228  
 <211> 24  
 <212> DNA  
 <213> Homo sapiens  
 <400> 228  
 agccttcgtg cccctaacca agtg 24  
 <210> 229  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens  
 <400> 229  
 ctgtgggcat tggggctcag g 21  
 <210> 230  
 <211> 20  
 <212> DNA  
 <213> Homo sapiens  
 <400> 230  
 gccccagtg gataccatg 20  
 <210> 231  
 <211> 21  
 <212> DNA  
 <213> Homo sapiens  
 <400> 231  
 atgctggagg ggacccacg g 21  
 <210> 232  
 <211> 3671  
 <212> DNA  
 <213> Homo sapiens  
 <220>  
 <221> CDS  
 <222> (418)...(3381)  
 <400> 232  
 tcctgggttca agcgattctc tggcctcagc ctccccgagta gctgggatta caggcatgct 60  
 ccaccaagcc cgggtaattt tgtattttta atagagacgg ggttttgcca tgttggtcag 120  
 gctgggtctcg aactcctgac ctcaggtgat ctgcccacct tggcctccca acgtgctgag 180  
 attacaggca tgagtcactg tgcccggcag agatgggtcta attcatatga aagaactctg 240



63/68

|   |            |            |            |             |             |     |
|---|------------|------------|------------|-------------|-------------|-----|
| aaaaaagtag  | aaagtgattt | tctaaaataa | ggtacaaata | attaatgtaa  | gcataatcac  | 300 |
| ctaaccttgt  | ggaatttttt | ttttttgaga | agcaaattgc | aaattttgtga | tagatctaaa  | 360 |
| ggagattgac  | taagagggtg | accatctgga | aatgacgtca | tgtgagaatg  | gttaaag atg | 420 |
|   |            |            |            |             | Met         |     |
|   |            |            |            |             | 1           |     |
| ctc ggg aga ttg agc cta gag aaa gga aga ttt gtg aac cca gga ggc | 468        |            |            |             |             |     |
| Leu Gly Arg Leu Ser Leu Glu Lys Gly Arg Phe Val Asn Pro Gly Gly |            |            |            |             |             |     |
| 5 10 15   |            |            |            |             |             |     |
| aga ggt aga gat cca gga gag ggc ggc gtg atg gat gac aag agt gaa | 516        |            |            |             |             |     |
| Arg Gly Arg Asp Pro Gly Glu Gly Gly Val Met Asp Asp Lys Ser Glu |            |            |            |             |             |     |
| 20 25 30  |            |            |            |             |             |     |
| gat tcc atg tcc gtc tcc acc ttg agc ttc ggt gtg aac aga ccc acg | 564        |            |            |             |             |     |
| Asp Ser Met Ser Val Ser Thr Leu Ser Phe Gly Val Asn Arg Pro Thr |            |            |            |             |             |     |
| 35 40 45  |            |            |            |             |             |     |
| att tcc tgc ata ttc gac tat ggg aac cgc tac cat cta cgc tgc tac | 612        |            |            |             |             |     |
| Ile Ser Cys Ile Phe Asp Tyr Gly Asn Arg Tyr His Leu Arg Cys Tyr |            |            |            |             |             |     |
| 50 55 60 65   |            |            |            |             |             |     |
| atg tac cag gcc cgg gac ctg gct gcg atg gac aag gac tct ttt tct | 660        |            |            |             |             |     |
| Met Tyr Gln Ala Arg Asp Leu Ala Ala Met Asp Lys Asp Ser Phe Ser |            |            |            |             |             |     |
| 70 75 80  |            |            |            |             |             |     |
| gat ccc tat gcc atc gtc tcc ttc ctg cac cag agc cag aag acg gtg | 708        |            |            |             |             |     |
| Asp Pro Tyr Ala Ile Val Ser Phe Leu His Gln Ser Gln Lys Thr Val |            |            |            |             |             |     |
| 85 90 95  |            |            |            |             |             |     |
| gtg gtg aag aac acc ctt aac ccc acc tgg gac cag acg ctc atc ttc | 756        |            |            |             |             |     |
| Val Val Lys Asn Thr Leu Asn Pro Thr Trp Asp Gln Thr Leu Ile Phe |            |            |            |             |             |     |
| 100 105 110   |            |            |            |             |             |     |
| tac gag atc gag atc ttt ggc gag ccg gcc aca gtt gct gag caa ccg | 804        |            |            |             |             |     |
| Tyr Glu Ile Glu Ile Phe Gly Glu Pro Ala Thr Val Ala Glu Gln Pro |            |            |            |             |             |     |
| 115 120 125   |            |            |            |             |             |     |
| ccc agc att gtg gtg gag ctg tac gac cat gac act tat ggt gca gac | 852        |            |            |             |             |     |
| Pro Ser Ile Val Val Glu Leu Tyr Asp His Asp Thr Tyr Gly Ala Asp |            |            |            |             |             |     |
| 130 135 140 145   |            |            |            |             |             |     |
| gag ttt atg ggt cgc tgc atc tgt caa ccg agt ctg gaa cgg atg cca | 900        |            |            |             |             |     |
| Glu Phe Met Gly Arg Cys Ile Cys Gln Pro Ser Leu Glu Arg Met Pro |            |            |            |             |             |     |
| 150 155 160   |            |            |            |             |             |     |
| cgg ctg gcc tgg ttc cca ctg acg agg ggc agc cag ccg tcg ggg gag | 948        |            |            |             |             |     |
| Arg Leu Ala Trp Phe Pro Leu Thr Arg Gly Ser Gln Pro Ser Gly Glu |            |            |            |             |             |     |
| 165 170 175   |            |            |            |             |             |     |
| ctg ctg gcc tct ttt gag ctg atc cag aga gag aag ccg gcc atc cac | 996        |            |            |             |             |     |
| Leu Leu Ala Ser Phe Glu Leu Ile Gln Arg Glu Lys Pro Ala Ile His |            |            |            |             |             |     |
| 180 185 190   |            |            |            |             |             |     |
| cat att cct ggt ttt gag gtg cag gag aca tca agg atc ctg gat gag | 1044       |            |            |             |             |     |
| His Ile Pro Gly Phe Glu Val Gln Glu Thr Ser Arg Ile Leu Asp Glu |            |            |            |             |             |     |
| 195 200 205   |            |            |            |             |             |     |
| tct gag gac aca gac ctg ccc tac cca cca ccc cag agg gag gcc aac | 1092       |            |            |             |             |     |
| Ser Glu Asp Thr Asp Leu Pro Tyr Pro Pro Pro Gln Arg Glu Ala Asn |            |            |            |             |             |     |
| 210 215 220 225   |            |            |            |             |             |     |
| atc tac atg gtt cct cag aac atc aag cca gcg ctc cag cgt acc gcc | 1140       |            |            |             |             |     |
| Ile Tyr Met Val Pro Gln Asn Ile Lys Pro Ala Leu Gln Arg Thr Ala |            |            |            |             |             |     |
| 230 235 240   |            |            |            |             |             |     |

64/68

|   |      |
|---|------|
| atc gag atc ctg gca tgg ggc ctg cgg aac atg aag agt tac cag ctg<br>Ile Glu Ile Leu Ala Trp Gly Leu Arg Asn Met Lys Ser Tyr Gln Leu<br>245 250 255     | 1188 |
| gcc aac atc tcc tcc ccc agc ctc gtg gta gag tgt ggg ggc cag acg<br>Ala Asn Ile Ser Ser Pro Ser Leu Val Val Glu Cys Gly Gly Gln Thr<br>260 265 270     | 1236 |
| gtg cag tcc tgt gtc atc agg aac ctc cgg aag aac ccc aac ttt gac<br>Val Gln Ser Cys Val Ile Arg Asn Leu Arg Lys Asn Pro Asn Phe Asp<br>275 280 285     | 1284 |
| atc tgc acc ctc ttc atg gaa gtg atg ctg ccc agg gag gag ctc tac<br>Ile Cys Thr Leu Phe Met Glu Val Met Leu Pro Arg Glu Glu Leu Tyr<br>290 295 300 305 | 1332 |
| tgc ccc ccc atc acc gtc aag gtc atc gat aac cgc cag ttt ggc cgc<br>Cys Pro Pro Ile Thr Val Lys Val Ile Asp Asn Arg Gln Phe Gly Arg<br>310 315 320     | 1380 |
| cgg cct gtg gtg ggc cag tgt acc atc cgc tcc ctg gag agc ttc ctg<br>Arg Pro Val Val Gly Gln Cys Thr Ile Arg Ser Leu Glu Ser Phe Leu<br>325 330 335     | 1428 |
| tgt gac ccc tac tcg gcg gag agt cca tcc cca cag ggt ggc cca gac<br>Cys Asp Pro Tyr Ser Ala Glu Ser Pro Ser Pro Gln Gly Gly Pro Asp<br>340 345 350     | 1476 |
| gat gtg agc cta ctc agt cct ggg gaa gac gtg ctc atc gac att gat<br>Asp Val Ser Leu Leu Ser Pro Gly Glu Asp Val Leu Ile Asp Ile Asp<br>355 360 365     | 1524 |
| gac aag gag ccc ctc atc ccc atc cag gag gaa gag ttc atc gat tgg<br>Asp Lys Glu Pro Leu Ile Pro Ile Gln Glu Glu Glu Phe Ile Asp Trp<br>370 375 380 385 | 1572 |
| tgg agc aaa ttc ttt gcc tcc ata ggg gag agg gaa aag tgc ggc tcc<br>Trp Ser Lys Phe Phe Ala Ser Ile Gly Glu Arg Glu Lys Cys Gly Ser<br>390 395 400     | 1620 |
| tac ctg gag aag gat ttt gac acc ctg aag gtc tat gac aca cag ctg<br>Tyr Leu Glu Lys Asp Phe Asp Thr Leu Lys Val Tyr Asp Thr Gln Leu<br>405 410 415     | 1668 |
| gag aat gtg gag gcc ttt gag ggc ctg tct gac ttt tgt aac acc ttc<br>Glu Asn Val Glu Ala Phe Glu Gly Leu Ser Asp Phe Cys Asn Thr Phe<br>420 425 430     | 1716 |
| aag ctg tac cgg ggc aag acg cag gag gag aca gaa gat cca tct gtg<br>Lys Leu Tyr Arg Gly Lys Thr Gln Glu Glu Thr Glu Asp Pro Ser Val<br>435 440 445     | 1764 |
| att ggt gaa ttt aag ggc ctc ttc aaa att tat ccc ctc cca gaa gac<br>Ile Gly Glu Phe Lys Lys Leu Phe Lys Ile Tyr Pro Leu Pro Glu Asp<br>450 455 460 465 | 1812 |
| cca gcc atc ccc atg ccc cca aga cag ttc cac cag ctg gcc gcc cag<br>Pro Ala Ile Pro Met Pro Pro Arg Gln Phe His Gln Leu Ala Ala Gln<br>470 475 480     | 1860 |
| gga ccc cag gag tgc ttg gtc cgt atc tac att gtc cga gca ttt ggc<br>Gly Pro Gln Glu Cys Leu Val Arg Ile Tyr Ile Val Arg Ala Phe Gly<br>485 490 495     | 1908 |
| ctg cag ccc aag gac ccc aat gga aag tgt gat cct tac atc aag atc<br>Leu Gln Pro Lys Asp Pro Asn Gly Lys Cys Asp Pro Tyr Ile Lys Ile<br>500 505 510     | 1956 |

65/68

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |      |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|
| tcc | ata | ggg | aag | aaa | tca | gtg | agt | gac | cag | gat | aac | tac | atc | ccc | tgc | 2004 |
| Ser | Ile | Gly | Lys | Lys | Ser | Val | Ser | Asp | Gln | Asp | Asn | Tyr | Ile | Pro | Cys |      |
| 515 |     |     |     |     |     | 520 |     |     |     |     | 525 |     |     |     |     |      |
| acg | ctg | gag | ccc | gta | ttt | gga | aag | atg | ttc | gag | ctg | acc | tgc | act | ctg | 2052 |
| Thr | Leu | Glu | Pro | Val | Phe | Gly | Lys | Met | Phe | Glu | Leu | Thr | Cys | Thr | Leu |      |
| 530 |     |     |     |     | 535 |     |     |     |     | 540 |     |     |     |     | 545 |      |
| cct | ctg | gag | aag | gac | cta | aag | atc | act | ctc | tat | gac | tat | gac | ctc | ctc | 2100 |
| Pro | Leu | Glu | Lys | Asp | Leu | Lys | Ile | Thr | Leu | Tyr | Asp | Tyr | Asp | Leu | Leu |      |
|     |     |     |     | 550 |     |     |     |     | 555 |     |     |     |     | 560 |     |      |
| tcc | aag | gac | gaa | aag | atc | ggt | gag | acg | gtc | gtc | gac | ctg | gag | aac | agg | 2148 |
| Ser | Lys | Asp | Glu | Lys | Ile | Gly | Glu | Thr | Val | Val | Asp | Leu | Glu | Asn | Arg |      |
|     |     |     | 565 |     |     |     |     | 570 |     |     |     |     | 575 |     |     |      |
| ctg | ctg | tcc | aag | ttt | ggg | gct | cgc | tgt | gga | ctc | cca | cag | acc | tac | tgt | 2196 |
| Leu | Leu | Ser | Lys | Phe | Gly | Ala | Arg | Cys | Gly | Leu | Pro | Gln | Thr | Tyr | Cys |      |
|     |     | 580 |     |     |     |     | 585 |     |     |     |     | 590 |     |     |     |      |
| gtc | tct | gga | ccg | aac | cag | tgg | cgg | gac | cag | ctc | cgc | ccc | tcc | cag | ctc | 2244 |
| Val | Ser | Gly | Pro | Asn | Gln | Trp | Arg | Asp | Gln | Leu | Arg | Pro | Ser | Gln | Leu |      |
|     | 595 |     |     |     |     | 600 |     |     |     |     | 605 |     |     |     |     |      |
| ctc | cac | ctc | ttc | tgc | cag | cag | cat | aga | gtc | aag | gca | cct | gtg | tac | cgg | 2292 |
| Leu | His | Leu | Phe | Cys | Gln | Gln | His | Arg | Val | Lys | Ala | Pro | Val | Tyr | Arg |      |
| 610 |     |     |     |     | 615 |     |     |     |     | 620 |     |     |     |     | 625 |      |
| aca | gac | cgt | gta | atg | ttt | cag | gat | aaa | gaa | tat | tcc | att | gaa | gag | ata | 2340 |
| Thr | Asp | Arg | Val | Met | Phe | Gln | Asp | Lys | Glu | Tyr | Ser | Ile | Glu | Glu | Ile |      |
|     |     |     |     | 630 |     |     |     |     | 635 |     |     |     |     | 640 |     |      |
| gag | gct | ggc | agg | atc | cca | aac | cca | cac | ctg | ggc | cca | gtg | gag | gag | cgt | 2388 |
| Glu | Ala | Gly | Arg | Ile | Pro | Asn | Pro | His | Leu | Gly | Pro | Val | Glu | Glu | Arg |      |
|     |     |     | 645 |     |     |     |     | 650 |     |     |     |     | 655 |     |     |      |
| ctg | gct | ctg | cat | gtg | ctt | cag | cag | cag | ggc | ctg | gtc | ccg | gag | cac | gtg | 2436 |
| Leu | Ala | Leu | His | Val | Leu | Gln | Gln | Gln | Gly | Leu | Val | Pro | Glu | His | Val |      |
|     |     | 660 |     |     |     |     | 665 |     |     |     |     | 670 |     |     |     |      |
| gag | tca | cgg | ccc | ctc | tac | agc | ccc | ctg | cag | cca | gac | atc | gag | cag | ggg | 2484 |
| Glu | Ser | Arg | Pro | Leu | Tyr | Ser | Pro | Leu | Gln | Pro | Asp | Ile | Glu | Gln | Gly |      |
|     | 675 |     |     |     |     | 680 |     |     |     |     | 685 |     |     |     |     |      |
| aag | ctg | cag | atg | tgg | gtc | gac | cta | ttt | ccg | aag | gcc | ctg | ggg | cgg | cct | 2532 |
| Lys | Leu | Gln | Met | Trp | Val | Asp | Leu | Phe | Pro | Lys | Ala | Leu | Gly | Arg | Pro |      |
| 690 |     |     |     |     | 695 |     |     |     |     | 700 |     |     |     |     | 705 |      |
| gga | cct | ccc | ttc | aac | atc | acc | cca | cgg | aga | gcc | aga | agg | ttt | ttc | ctg | 2580 |
| Gly | Pro | Pro | Phe | Asn | Ile | Thr | Pro | Arg | Arg | Ala | Arg | Arg | Phe | Phe | Leu |      |
|     |     |     | 710 |     |     |     |     |     | 715 |     |     |     |     | 720 |     |      |
| cgt | tgt | att | atc | tgg | aat | acc | aga | gat | gtg | atc | ctg | gat | gac | ctg | agc | 2628 |
| Arg | Cys | Ile | Ile | Trp | Asn | Thr | Arg | Asp | Val | Ile | Leu | Asp | Asp | Leu | Ser |      |
|     |     |     | 725 |     |     |     |     | 730 |     |     |     |     | 735 |     |     |      |
| ctc | acg | ggg | gag | aag | atg | agc | gac | att | tat | gtg | aaa | ggt | tgg | atg | att | 2676 |
| Leu | Thr | Gly | Glu | Lys | Met | Ser | Asp | Ile | Tyr | Val | Lys | Gly | Trp | Met | Ile |      |
|     |     | 740 |     |     |     |     | 745 |     |     |     |     | 750 |     |     |     |      |
| ggc | ttt | gaa | gaa | cac | aag | caa | aag | aca | gac | gtg | cat | tat | cgt | tcc | ctg | 2724 |
| Gly | Phe | Glu | Glu | His | Lys | Gln | Lys | Thr | Asp | Val | His | Tyr | Arg | Ser | Leu |      |
|     | 755 |     |     |     |     | 760 |     |     |     |     | 765 |     |     |     |     |      |

66/68

|   |      |
|---|------|
| gga ggt gaa ggc aac ttc aac tgg agg ttc att ttc ccc ttc gac tac     | 2772 |
| Gly Gly Glu Gly Asn Phe Asn Trp Arg Phe Ile Phe Pro Phe Asp Tyr     |      |
| 770 775 780 785   |      |
| ctg cca gct gag caa gtc tgt acc att gcc aag aag gat gcc ttc tgg     | 2820 |
| Leu Pro Ala Glu Val Cys Thr Ile Ala Lys Lys Asp Ala Phe Trp         |      |
| 790 795 800   |      |
| agg ctg gac aag act gag agc aaa atc cca gca cga gtg gtg ttc cag     | 2868 |
| Arg Leu Asp Lys Thr Glu Ser Lys Ile Pro Ala Arg Val Val Phe Gln     |      |
| 805 810 815   |      |
| atc tgg gac aat gac aag ttc tcc ttt gat gat ttt ctg ggc tcc ctg     | 2916 |
| Ile Trp Asp Asn Asp Lys Phe Ser Phe Asp Asp Phe Leu Gly Ser Leu     |      |
| 820 825 830   |      |
| cag ctc gat ctc aac cgc atg ccc aag cca gcc aag aca gcc aag aag     | 2964 |
| Gln Leu Asp Leu Asn Arg Met Pro Lys Pro Ala Lys Thr Ala Lys Lys     |      |
| 835 840 845   |      |
| tgc tcc ttg gac cag ctg gat gat gct ttc cac cca gaa tgg ttt gtg     | 3012 |
| Cys Ser Leu Asp Gln Leu Asp Asp Ala Phe His Pro Glu Trp Phe Val     |      |
| 850 855 860 865   |      |
| tcc ctt ttt gag cag aaa aca gtg aag ggc tgg tgg ccc tgt gta gca     | 3060 |
| Ser Leu Phe Glu Gln Lys Thr Val Lys Gly Trp Trp Pro Cys Val Ala     |      |
| 870 875 880   |      |
| gaa gag ggt gag aag aaa ata ctg gcg ggc aag ctg gaa atg acc ttg     | 3108 |
| Glu Glu Gly Glu Lys Lys Ile Leu Ala Gly Lys Leu Glu Met Thr Leu     |      |
| 885 890 895   |      |
| gag att gta gca gag agt gag cat gag gag cgg cct gct ggc cag ggc     | 3156 |
| Glu Ile Val Ala Glu Ser Glu His Glu Glu Arg Pro Ala Gly Gln Gly     |      |
| 900 905 910   |      |
| cgg gat gag ccc aac atg aac cct aag ctt gag gac cca agg cgc ccc     | 3204 |
| Arg Asp Glu Pro Asn Met Asn Pro Lys Leu Glu Asp Pro Arg Arg Pro     |      |
| 915 920 925   |      |
| gac acc tcc ttc ctg tgg ttt acc tcc cca tac aag acc atg aag ttc     | 3252 |
| Asp Thr Ser Phe Leu Trp Phe Thr Ser Pro Tyr Lys Thr Met Lys Phe     |      |
| 930 935 940 945   |      |
| atc ctg tgg cgg cgt ttc cgg tgg gcc atc atc ctc ttc atc atc ctc     | 3300 |
| Ile Leu Trp Arg Arg Phe Arg Trp Ala Ile Ile Leu Phe Ile Ile Leu     |      |
| 950 955 960   |      |
| ttc atc ctg ctg ctg ttc ctg gcc atc ttc atc tac gcc ttc ccg aac     | 3348 |
| Phe Ile Leu Leu Phe Leu Ala Ile Phe Ile Tyr Ala Phe Pro Asn         |      |
| 965 970 975   |      |
| tat gct gcc atg aag ctg gtg aag ccc ttc agc tgaggactct cctgccctgt   | 3401 |
| Tyr Ala Ala Met Lys Leu Val Lys Pro Phe Ser                         |      |
| 980 985   |      |
| agaagggggcc gtgggggtccc ctccagcatg ggactggcct gcctcctccg cccagctcgg | 3461 |
| cgagctcctc cagacctcct aggcctgatt gtccctgccag ggtgggcaga cagacagatg  | 3521 |
| gaccggccca cactcccaga gttgctaaca tggagctctg agatcacccc acttccatca   | 3581 |
| tttccttctc ccccaaccca acgctttttt ggatcagctc agacatatatt cagtataaaa  | 3641 |
| cagttggaac cacaaaaaaa aaaaaaaaaa                                    | 3671 |

<210> 233  
 <211> 988  
 <212> PRT  
 <213> Homo sapiens

67/68

<400> 233  
 Met Leu Gly Arg Leu Ser Leu Glu Lys Gly Arg Phe Val Asn Pro Gly  
 1 5 10 15  
 Gly Arg Gly Arg Asp Pro Gly Glu Gly Gly Val Met Asp Asp Lys Ser  
 20 25 30  
 Glu Asp Ser Met Ser Val Ser Thr Leu Ser Phe Gly Val Asn Arg Pro  
 35 40 45  
 Thr Ile Ser Cys Ile Phe Asp Tyr Gly Asn Arg Tyr His Leu Arg Cys  
 50 55 60  
 Tyr Met Tyr Gln Ala Arg Asp Leu Ala Ala Met Asp Lys Asp Ser Phe  
 65 70 75 80  
 Ser Asp Pro Tyr Ala Ile Val Ser Phe Leu His Gln Ser Gln Lys Thr  
 85 90 95  
 Val Val Val Lys Asn Thr Leu Asn Pro Thr Trp Asp Gln Thr Leu Ile  
 100 105 110  
 Phe Tyr Glu Ile Glu Ile Phe Gly Glu Pro Ala Thr Val Ala Glu Gln  
 115 120 125  
 Pro Pro Ser Ile Val Val Glu Leu Tyr Asp His Asp Thr Tyr Gly Ala  
 130 135 140  
 Asp Glu Phe Met Gly Arg Cys Ile Cys Gln Pro Ser Leu Glu Arg Met  
 145 150 155 160  
 Pro Arg Leu Ala Trp Phe Pro Leu Thr Arg Gly Ser Gln Pro Ser Gly  
 165 170 175  
 Glu Leu Leu Ala Ser Phe Glu Leu Ile Gln Arg Glu Lys Pro Ala Ile  
 180 185 190  
 His His Ile Pro Gly Phe Glu Val Gln Glu Thr Ser Arg Ile Leu Asp  
 195 200 205  
 Glu Ser Glu Asp Thr Asp Leu Pro Tyr Pro Pro Pro Gln Arg Glu Ala  
 210 215 220  
 Asn Ile Tyr Met Val Pro Gln Asn Ile Lys Pro Ala Leu Gln Arg Thr  
 225 230 235 240  
 Ala Ile Glu Ile Leu Ala Trp Gly Leu Arg Asn Met Lys Ser Tyr Gln  
 245 250 255  
 Leu Ala Asn Ile Ser Ser Pro Ser Leu Val Val Glu Cys Gly Gly Gln  
 260 265 270  
 Thr Val Gln Ser Cys Val Ile Arg Asn Leu Arg Lys Asn Pro Asn Phe  
 275 280 285  
 Asp Ile Cys Thr Leu Phe Met Glu Val Met Leu Pro Arg Glu Glu Leu  
 290 295 300  
 Tyr Cys Pro Pro Ile Thr Val Lys Val Ile Asp Asn Arg Gln Phe Gly  
 305 310 315 320  
 Arg Arg Pro Val Val Gly Gln Cys Thr Ile Arg Ser Leu Glu Ser Phe  
 325 330 335  
 Leu Cys Asp Pro Tyr Ser Ala Glu Ser Pro Ser Pro Gln Gly Gly Pro  
 340 345 350  
 Asp Asp Val Ser Leu Leu Ser Pro Gly Glu Asp Val Leu Ile Asp Ile  
 355 360 365  
 Asp Asp Lys Glu Pro Leu Ile Pro Ile Gln Glu Glu Glu Phe Ile Asp  
 370 375 380  
 Trp Trp Ser Lys Phe Phe Ala Ser Ile Gly Glu Arg Glu Lys Cys Gly  
 385 390 395 400  
 Ser Tyr Leu Glu Lys Asp Phe Asp Thr Leu Lys Val Tyr Asp Thr Gln  
 405 410 415  
 Leu Glu Asn Val Glu Ala Phe Glu Gly Leu Ser Asp Phe Cys Asn Thr  
 420 425 430  
 Phe Lys Leu Tyr Arg Gly Lys Thr Gln Glu Glu Thr Glu Asp Pro Ser  
 435 440 445  
 Val Ile Gly Glu Phe Lys Gly Leu Phe Lys Ile Tyr Pro Leu Pro Glu  
 450 455 460  
 Asp Pro Ala Ile Pro Met Pro Pro Arg Gln Phe His Gln Leu Ala Ala  
 465 470 475 480  
 Gln Gly Pro Gln Glu Cys Leu Val Arg Ile Tyr Ile Val Arg Ala Phe  
 485 490 495  
 Gly Leu Gln Pro Lys Asp Pro Asn Gly Lys Cys Asp Pro Tyr Ile Lys  
 500 505 510  
 Ile Ser Ile Gly Lys Lys Ser Val Ser Asp Gln Asp Asn Tyr Ile Pro  
 515 520 525

68/68

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Cys | Thr | Leu | Glu | Pro | Val | Phe | Gly | Lys | Met | Phe | Glu | Leu | Thr | Cys | Thr |
|     | 530 |     |     |     |     | 535 |     |     |     |     | 540 |     |     |     |     |
| Leu | Pro | Leu | Glu | Lys | Asp | Leu | Lys | Ile | Thr | Leu | Tyr | Asp | Tyr | Asp | Leu |
| 545 |     |     |     |     | 550 |     |     |     |     | 555 |     |     |     |     | 560 |
| Leu | Ser | Lys | Asp | Glu | Lys | Ile | Gly | Glu | Thr | Val | Val | Asp | Leu | Glu | Asn |
|     |     |     |     | 565 |     |     |     |     | 570 |     |     |     |     | 575 |     |
| Arg | Leu | Leu | Ser | Lys | Phe | Gly | Ala | Arg | Cys | Gly | Leu | Pro | Gln | Thr | Tyr |
|     |     |     | 580 |     |     |     |     | 585 |     |     |     |     | 590 |     |     |
| Cys | Val | Ser | Gly | Pro | Asn | Gln | Trp | Arg | Asp | Gln | Leu | Arg | Pro | Ser | Gln |
|     |     | 595 |     |     |     |     | 600 |     |     |     |     |     | 605 |     |     |
| Leu | Leu | His | Leu | Phe | Cys | Gln | Gln | His | Arg | Val | Lys | Ala | Pro | Val | Tyr |
|     | 610 |     |     |     |     | 615 |     |     |     |     | 620 |     |     |     |     |
| Arg | Thr | Asp | Arg | Val | Met | Phe | Gln | Asp | Lys | Glu | Tyr | Ser | Ile | Glu | Glu |
| 625 |     |     |     |     | 630 |     |     |     |     | 635 |     |     |     |     | 640 |
| Ile | Glu | Ala | Gly | Arg | Ile | Pro | Asn | Pro | His | Leu | Gly | Pro | Val | Glu | Glu |
|     |     |     |     | 645 |     |     |     |     | 650 |     |     |     |     | 655 |     |
| Arg | Leu | Ala | Leu | His | Val | Leu | Gln | Gln | Gln | Gly | Leu | Val | Pro | Glu | His |
|     |     |     | 660 |     |     |     |     | 665 |     |     |     |     |     | 670 |     |
| Val | Glu | Ser | Arg | Pro | Leu | Tyr | Ser | Pro | Leu | Gln | Pro | Asp | Ile | Glu | Gln |
|     |     | 675 |     |     |     |     | 680 |     |     |     |     |     | 685 |     |     |
| Gly | Lys | Leu | Gln | Met | Trp | Val | Asp | Leu | Phe | Pro | Lys | Ala | Leu | Gly | Arg |
|     | 690 |     |     |     |     | 695 |     |     |     |     | 700 |     |     |     |     |
| Pro | Gly | Pro | Pro | Phe | Asn | Ile | Thr | Pro | Arg | Arg | Ala | Arg | Arg | Phe | Phe |
| 705 |     |     |     |     | 710 |     |     |     |     | 715 |     |     |     |     | 720 |
| Leu | Arg | Cys | Ile | Ile | Trp | Asn | Thr | Arg | Asp | Val | Ile | Leu | Asp | Asp | Leu |
|     |     |     |     | 725 |     |     |     |     | 730 |     |     |     |     | 735 |     |
| Ser | Leu | Thr | Gly | Glu | Lys | Met | Ser | Asp | Ile | Tyr | Val | Lys | Gly | Trp | Met |
|     |     |     | 740 |     |     |     |     | 745 |     |     |     |     | 750 |     |     |
| Ile | Gly | Phe | Glu | Glu | His | Lys | Gln | Lys | Thr | Asp | Val | His | Tyr | Arg | Ser |
|     |     | 755 |     |     |     |     | 760 |     |     |     |     | 765 |     |     |     |
| Leu | Gly | Gly | Glu | Gly | Asn | Phe | Asn | Trp | Arg | Phe | Ile | Phe | Pro | Phe | Asp |
|     | 770 |     |     |     |     | 775 |     |     |     |     | 780 |     |     |     |     |
| Tyr | Leu | Pro | Ala | Glu | Gln | Val | Cys | Thr | Ile | Ala | Lys | Lys | Asp | Ala | Phe |
| 785 |     |     |     |     | 790 |     |     |     |     | 795 |     |     |     |     | 800 |
| Trp | Arg | Leu | Asp | Lys | Thr | Glu | Ser | Lys | Ile | Pro | Ala | Arg | Val | Val | Phe |
|     |     |     |     | 805 |     |     |     |     | 810 |     |     |     |     | 815 |     |
| Gln | Ile | Trp | Asp | Asn | Asp | Lys | Phe | Ser | Phe | Asp | Asp | Phe | Leu | Gly | Ser |
|     |     |     | 820 |     |     |     |     | 825 |     |     |     |     | 830 |     |     |
| Leu | Gln | Leu | Asp | Leu | Asn | Arg | Met | Pro | Lys | Pro | Ala | Lys | Thr | Ala | Lys |
|     |     | 835 |     |     |     |     | 840 |     |     |     |     |     | 845 |     |     |
| Lys | Cys | Ser | Leu | Asp | Gln | Leu | Asp | Asp | Ala | Phe | His | Pro | Glu | Trp | Phe |
|     | 850 |     |     |     |     | 855 |     |     |     |     | 860 |     |     |     |     |
| Val | Ser | Leu | Phe | Glu | Gln | Lys | Thr | Val | Lys | Gly | Trp | Trp | Pro | Cys | Val |
| 865 |     |     |     |     | 870 |     |     |     |     | 875 |     |     |     |     | 880 |
| Ala | Glu | Glu | Gly | Glu | Lys | Lys | Ile | Leu | Ala | Gly | Lys | Leu | Glu | Met | Thr |
|     |     |     |     | 885 |     |     |     |     | 890 |     |     |     |     | 895 |     |
| Leu | Glu | Ile | Val | Ala | Glu | Ser | Glu | His | Glu | Glu | Arg | Pro | Ala | Gly | Gln |
|     |     |     | 900 |     |     |     |     | 905 |     |     |     |     | 910 |     |     |
| Gly | Arg | Asp | Glu | Pro | Asn | Met | Asn | Pro | Lys | Leu | Glu | Asp | Pro | Arg | Arg |
|     |     | 915 |     |     |     |     | 920 |     |     |     |     |     | 925 |     |     |
| Pro | Asp | Thr | Ser | Phe | Leu | Trp | Phe | Thr | Ser | Pro | Tyr | Lys | Thr | Met | Lys |
|     |     | 930 |     |     |     | 935 |     |     |     |     | 940 |     |     |     |     |
| Phe | Ile | Leu | Trp | Arg | Arg | Phe | Arg | Trp | Ala | Ile | Ile | Leu | Phe | Ile | Ile |
| 945 |     |     |     |     | 950 |     |     |     |     | 955 |     |     |     |     | 960 |
| Leu | Phe | Ile | Leu | Leu | Phe | Leu | Ala | Ile | Phe | Ile | Tyr | Ala | Phe | Phe | Pro |
|     |     |     |     | 965 |     |     |     |     | 970 |     |     |     |     | 975 |     |
| Asn | Tyr | Ala | Ala | Met | Lys | Leu | Val | Lys | Pro | Phe | Ser |     |     |     |     |
|     |     |     | 980 |     |     |     |     | 985 |     |     |     |     |     |     |     |

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US99/19395

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(7) : C12N 15/11, 15/00; C07K 16/00

US CL : 536/23.1, 435/440, 530/387.1

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1, 435/440, 530/387.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

BIOSIS, CAPLUS, EMBASE, EMBASE, LIFESCI, MEDLINE, SCISEARCH, TOXLIT

Search Terms: dysferlin, lgmd2b

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages  | Relevant to claim No. |
|-----------|---|-----------------------|
| X         | WEILER et al. Limb-girdle muscular dystrophy and Myoshi Myopathy in an aboriginal Canadian kindred map to LGMD2B and segregate with the same haplotype. American Journal of Human Genetics. October 1996, Vol.59, pages 872-878, especially page 873. | 32,35                 |
| X         | KOENIG et al. Complete cloning of the Duchenne Muscular Dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. Cell. 31 July 1987, Vol. 50, pages 509-517, especially pages 511-513.           | 32-33,36              |

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

|   |  |
|---|--|
| * Special categories of cited documents:  | *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention  |
| *A* document defining the general state of the art which is not considered to be of particular relevance  | *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone   |
| *B* earlier document published on or after the international filing date  | *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |
| *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | *Z* document member of the same patent family  |
| *O* document referring to an oral disclosure, use, exhibition or other means  |  |
| *P* document published prior to the international filing date but later than the priority date claimed  |  |

|   |  |
|---|--|
| Date of the actual completion of the international search<br>17 NOVEMBER 1999   | Date of mailing of the international search report<br><b>13 JAN 2000</b> |
| Name and mailing address of the ISA/US<br>Commissioner of Patents and Trademarks<br>Box PCT<br>Washington, D.C. 20231<br>Facsimile No. (703) 305-4242 | Authorized officer<br>Stephen Siu<br>Telephone No. (703) 308-0196        |

Form PCT/ISA/210 (second sheet)(July 1992)\*

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US99/19395

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category*         | Citation of document, with indication, where appropriate, of the relevant passages   | Relevant to claim No.         |
|-------------------|--|-------------------------------|
| X,P<br>---<br>Y,P | Database GenCore version 4.5, Compugen Ltd., No. AI128455, 'NCI-CGAP, National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index', Unpublished, 27 October 1998   | 1,6,12<br>-----<br>7,14,16    |
| X<br>---<br>Y     | Database GenCore version 4.5, Compugen Ltd., No. R41062, WAYE, M.M.Y. et al. 'Gene expression of adult human heart as revealed by random sequencing of cDNA library,' Miami Winter Biotechnol. Symp. Proc. 6,90 , 16 May 16, 1995. | 1, 6, 11-12<br>-----<br>7, 14 |
| X<br>---<br>Y     | Database GenCore version 4.5, Compugen Ltd., No. AA718275, Marra et al, 'The WashU-HHMI Mouse EST Project', Unpublished, 29 December 1997.   | 1, 6, 11-12<br>-----<br>7, 14 |
| Y                 | BASHIR et al. Genetic and physical mapping at the limb-girdle muscular dystrophy locus (LGMD2B) on chromosome 2p. Genomics. April 1996, Vol.33, pages 46-52, especially page 47.   | 32,36                         |
| Y                 | MOREIRA et al. The seventh form of autosomal recessive limb-girdle muscular dystrophy is mapped to 17q11-12. American Journal of Human Genetics. July 1997, Vol. 61, pages 151-159, entire document.                               | 32, 35                        |
| Y                 | Database GenCore version 4.5, Compugen Ltd., No. R76778, HILLIER et al., 'The WashU-Merck EST Project', Unpublished, 06 June 1995.   | 7, 14                         |
| A,E               | AHLBERG et al. Genetic Linkage of Welander Distal Myopathy to chromosome 2p13. Annals of Neurology. September 1999, Vol. 46, No.3, pages 399-404, especially page 400.   | 37, 39                        |
| A,E               | BITTNER et al. Dysferlin deletion in SJL mice (SJL-Dysf) defines a natural model for limb girdle muscular dystrophy 2B. Nature Genetics. October 1999, Vol. 23, pages 141-142, especially page 141.                                | 40                            |
| A,P               | BASHIR et al. A gene related to Caenorhabditis elegans spermatogenesis factor fer-1 is mutated in limb-girdle muscular dystrophy type 2B. Nature Genetics. September 1998, Vol 20, pages 37-42.                                    | 1-53                          |
| A,E               | Matsuda et al. Dysferlin is a surface membrane-associated protein that is absent in Miyoshi Myopathy. Neurology 22 September 1999, Vol. 53, No. 5, pages 1119-1122, especially pages 1119-1120.                                    | 40                            |



INTERNATIONAL SEARCH REPORT

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages  | Relevant to claim No. |
|-----------|---|-----------------------|
| A,P       | LIU et al. Dysferlin, a novel skeletal muscle gene, is mutated in Miyoshi Myopathy and limb girdle muscular dystrophy. Nature Genetics. September 1998, Vol. 20, pages 31-36. | 1-54                  |